Efficacy of Schoolwide Programs to Promote Social and Character Development and Reduce Problem Behavior in Elementary School Children

Efficacy of Schoolwide Programs to Promote Social and Character Development and Reduce Problem Behavior in Elementary School Children

Report From the Social and Character Development Research Program

OCTOBER 2010

Social and Character Development Research Consortium

Disclaimer

The findings and conclusions expressed in this report are the authors' and do not necessarily represent the opinions and positions of the Institute of Education Sciences, the U.S. Department of Education, or the Centers for Disease Control and Prevention.

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Research
Lynn Okagaki
Commissioner

October 2010

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 22207, Alexandria, VA 22304.

- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not available in your area, call 800-872-5327. Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.

- Fax your request to 703-605-6794 or order online at www.edpubs.gov.

The report is also available on the IES website at http://ncer.ed.gov.

On request, this publication is available in alternate formats, such as Braille, large print, or computer diskette. For more information, contact the Department's Alternate Format Center at 202-260-0852 or 202-260-0818.

Social and Character Development Research Consortium

Institute of Education Sciences (IES)
Caroline Ebanks
Edward C. Metz
Amy Silverman[1]
Elizabeth R. Albro

Division of Violence Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (CDC)
Tamara M. Haegerich[2]
Le'Roy Reese[3]
Corinne David-Ferdon

Mathematica Policy Research, Inc. (MPR)
Karen Needels
John A. Burghardt
Laura M. Kalb
Heather Koball
Peter Z. Schochet
Victor Battistich (University of Missouri-St. Louis)

Children's Institute
Deborah B. Johnson
Hugh F. Crean

New York University[4]
J. Lawrence Aber
Stephanie M. Jones[5] (Fordham University)
Joshua L. Brown[6]

Oregon State University[7]
Brian R. Flay
Carol G. Allred (Positive Action)
David L. DuBois (University of Illinois at Chicago)
Michael L. Berbaum (University of Illinois at Chicago)
Peter Ji (University of Illinois at Chicago)
Vanessa Brechling (University of Illinois at Chicago)

University at Buffalo, The State University of New York
William E. Pelham, Jr.
Greta M. Massetti[8]
Daniel A. Waschbusch

University of Maryland
Gary D. Gottfredson

University of North Carolina at Chapel Hill
Mark W. Fraser
Thomas W. Farmer[9]
Maeda J. Galinsky
Kim Dadisman

Vanderbilt University
Leonard Bickman
Ana Maria Brannan
Catherine Smith

---

[1] Amy Silverman is currently a clinical psychologist at the Jewish Family & Career Services of Atlanta, Inc.

[2] Tamara M. Haegerich served as IES Project Officer before joining the CDC.

[3] Le'Roy Reese is currently on the faculty at Morehouse School of Medicine.

[4] This award was originally made to Columbia University and was transferred to New York University (NYU) when the principal investigator, J. Lawrence Aber, accepted a position at NYU.

[5] Stephanie M. Jones is now on the faculty at Harvard University.

[6] Joshua L. Brown is now on the faculty at Fordham University.

[7] This award was originally made to the University of Illinois at Chicago, but was transferred to Oregon State University (OSU) when the principal investigator, Brian R. Flay, accepted a position at OSU.

[8] Greta M. Massetti is now at the Division of Violence Prevention, National Center for Injury Prevention and Control, CDC.

[9] Thomas W. Farmer was the principal investigator when the award was made to the University of North Carolina at Chapel Hill. During the project, he accepted a position at Pennsylvania State University, and Mark Fraser became the principal investigator.

This page is intentionally blank.

Acknowledgments

This page is intentionally blank.

Disclosure of Potential Conflicts of Interest

The Social and Character Development (SACD) Research Consortium consists of research teams (principal investigators and co-principal investigators from each grantee site); Institute of Education Sciences (IES) staff; Centers for Disease Control and Prevention (CDC) staff; and the evaluation contractor, Mathematica Policy Research, Inc. (MPR). Except for the three exceptions described in the paragraph below, the grantee research teams, IES staff, CDC staff, and contractor staff from MPR have no interests that could be affected by findings from the evaluation of the intervention programs that are described in this report.

Three of the SACD program research teams were led by and/or included personnel who had developed the intervention programs implemented at their respective research sites. Dr. William E. Pelham, Jr. and his colleagues developed the *Academic and Behavioral Competencies Program*, which was implemented in elementary schools in New York. The *Competence Support Program*, implemented in North Carolina schools, is made up of three distinct interventions (*Making Choices, Competence Enhancement Behavior Management,* and *Classroom Social Dynamics Training*). Dr. Mark W. Fraser and his colleagues developed the *Making Choices* program. Dr. Thomas W. Farmer and his colleagues developed the *Competence Enhancement Behavior Management* and the *Social Dynamics Training* programs. The *Positive Action* program was developed by Carol G. Allred, who is a member of the research team that implemented the program in schools in Illinois and is married to Dr. Brian R. Flay, the leader of that team.

All seven teams of researchers were selected to receive funding for their SACD research projects in a competitive grant application process. Each research team implemented its intervention and conducted site-specific analysis examining the effects of these interventions on student outcomes. MPR, the evaluation study contractor, conducted independent evaluations of all the intervention programs that were included in the SACD study. A data collection team from MPR independently collected the data for the evaluation presented in this report with the exception of the fidelity of implementation data, which were collected by the seven teams. The MPR data analysis team completed all descriptive and impact analyses. The developers/implementers of these interventions did not conduct the impact analyses that are summarized in this report.

This page is intentionally blank.

Glossary

4Rs—*The 4Rs Program (Reading, Writing, Respect, and Resolution)*

ABC—*Academic and Behavioral Competencies Program*

ADHD—attention deficit hyperactivity disorder

CDC—Centers for Disease Control and Prevention

combined-program—all seven programs together

CR—Child Report

CSP—*Competence Support Program*

DIR—Decision Information Resources, Inc., a subcontractor to Mathematica Policy Research, Inc. (MPR), the contractor selected to do the independent evaluation of the SACD programs

ES—effect size

HHS—U.S. Department of Health and Human Services

HLM—hierarchical linear model

ICC—intraclass correlation

IES—Institute of Education Sciences, U.S. Department of Education

LBW—*Love In a Big World*

leavers—students who left schools that were participating in the study before the study was completed

MANOVA—multivariate analysis of variance

MDES—minimum detectable effect size

middle childhood—the developmental period from 6 to 12 years of age

MPR—Mathematica Policy Research, Inc., the contractor selected to do the independent evaluation of the SACD programs

multiprogram evaluation—the evaluation of the impact of the seven SACD programs on students' social and character development. The evaluation looked at the average impact of all seven SACD programs combined and the average impact of each SACD program individually.

named activities—activities carried out in the context of a named SACD program to support students' social and character development

named program—a specific SACD program

new entrants—students who entered schools participating in the study after the study had begun

outcome domain—a set of outcomes grouped together because they measure a similar school activity, teacher instructional behavior, or student behavior

outcomes—what the SACD programs were expected to change regarding school activities, teacher instructional behavior, or student behavior

PA—*Positive Action*

PATHS—*Promoting Alternative Thinking Strategies*

PCR—Primary Caregiver Report

SACD—social and character development

SACD activities—activities to support students' social and character development

SACD goals—six specific goals defined by the SACD Research Program and promoted through the use of SACD activities to support students' social and character development. The goals are (1) violence prevention and peace promotion, (2) social and emotional development, (3) character education, (4) tolerance and diversity, (5) risk prevention and health promotion, and (6) civic responsibility and community service.

SACD program—a specific program that has the purpose of improving the social and character development of students

SACD Research Consortium—the researchers responsible for the SACD Research Program. These include staff at the Institute of Education Sciences (IES) in the U.S. Department of Education; the Division of Violence Prevention in the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (CDC); and Mathematica Policy Research, Inc. (MPR), the contractor, selected to do the independent evaluation of the SACD programs; as well as the seven research teams that received funding under cooperative agreements to evaluate one SACD program of their choosing under an experimental design.

SACD Research Program—a collaboration between the Institute of Education Sciences (IES) in the U.S. Department of Education and the Division of Violence Prevention in the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (CDC). Its purpose is to evaluate multiple universal, elementary school-based programs that target student social development and behavior outcomes.

SACD training—professional development for teachers on instructional methods to promote student social and character development

school-based program—a program delivered in a school setting

site—the set of schools, both treatment and control, specific to each individual SACD program evaluation

SS—*Second Step*

stayers—students who were in schools participating in the study for the entire study

targeted school-based programs—programs that are designed to address the needs of a subset of the children in a school

TRCS—Teacher Report on Classroom and School

TRS—Teacher Report on Student

universal programs—programs that are designed to address the needs of all children in a school

WWC—the What Works Clearinghouse

Contents

Contents

Contents

Contents

Contents

List of Tables

Table   **Page**

Executive Summary

Chapter 1. The Social and Character Development Multiprogram Evaluation

Table   **Page**

| **Table** | **Page** |

Table   **Page**

Chapter 7. The 4Rs Program (Reading, Writing, Respect, and Resolution)

Chapter 8. Second Step

List of Figures

Executive Summary

A variety of universal school-based programs designed to help elementary schools foster positive student behaviors, reduce negative behaviors, and, ultimately, improve academic performance are available; however, more evidence from rigorous evaluations is needed to better understand their effects. Such information is important because the development of social competencies during middle childhood has been linked to adjustment to schooling and academic success, while the failure to develop such competencies can lead to problem behavior that interferes with success in school (Bennett et al. 2003; Carlson et al. 1999; Farrington 1989; Fors, Crepaz, and Hayes 1999; Malecki and Elliot 2002; McCord et al. 2000; Najaka, Gottfredson, and Wilson 2001; O'Donnell, Hawkins, and Abbott 1995; Trzesniewski et al. 2006; Wentzel 1993).[1]

The Institute of Education Sciences (IES) and the Division of Violence Prevention in the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention (CDC) collaborated to conduct a rigorous impact evaluation of programs aimed at improving students' behavior. For this evaluation, such programs were termed Social and Character Development (SACD) programs.[2] Seven programs were evaluated, and all were coherent in that their activities were integrated and logically organized based on a theory of action (that differed among the programs), school-based in that they were implemented in the schools by school personnel, and universal in that they were to be implemented for all students in all elementary classrooms in a school.

This report provides the results from the evaluation of the seven SACD programs on one cohort of students as they moved from third through fifth grades starting in fall 2004 and ending in spring 2007.[3] The evaluation examined the effects on these students of the seven programs, together and separately, after 1, 2, and 3 school years and also estimated the impact on students' growth in social and character development over the 3 years. Chapter 1 discusses the evaluation of the programs when considered together and provides summary results for each program. Chapters 2 through 8 detail the findings for each of the programs individually. There are two appendixes: appendix A examines whether the addition of the smaller second cohort of students to the study affected the results, and appendix B contains additional technical information concerning the analyses.

---

[1] More information on the value of, evidence for, and theories behind these programs can be found in the second and third sections of chapter 1.

[2] Activities carried out by these programs in support of students' social and character development are termed SACD activities. The SACD evaluation examined SACD activities intended to promote six goals (termed SACD goals) and behavior management. The six SACD goals included (1) violence prevention and peace promotion, (2) social and emotional development, (3) character education, (4) tolerance and diversity, (5) risk prevention and health promotion, and (6) civic responsibility and community service.

[3] From 2005 to 2007, a smaller, second cohort of students was followed from third through fourth grades in a separate set of schools. This cohort is discussed in appendix A. The Executive Summary and chapters 1 through 8 describe the analysis of Cohort 1 only.

Study Design

Through a competitive application process that included a scientific peer review, seven research teams received funding under cooperative agreements to evaluate one SACD program of their choosing under an experimental design. Each research team recruited 10 to 14 schools[4] (for a total of 84), with half of the schools implementing one of the seven SACD programs for the 3 years of the study (the treatment group) and the other half continuing with their traditional SACD activities (the control group). Each team's schools were randomly assigned to the treatment or control group through a stratified sampling process. All but one school (a control school) participated in the study for the full 3 years. Table A lists the research teams, the SACD program each evaluated, key features of each program, and the number of treatment and control schools.[5]

Under a separate peer-reviewed competition, Mathematica Policy Research, Inc. (MPR) received a contract to carry out a multiprogram evaluation of the seven SACD programs using (1) standardized data collection for all sites, (2) a common set of descriptive measures on the types and level of school-based activities (i.e., SACD activities) that targeted social and behavioral outcomes at both the treatment and control schools, (3) a common set of outcome measures, and (4) a uniform statistical analysis.[6] The evaluation examined the impacts of all seven programs together, assessed the effect of each program separately to identify any contrasts with the overall findings across programs, and tested differences in effects on student subgroups.

Data were first collected from students, their primary caregivers, teachers, and principals in the fall of 2004 when students were starting third grade. Follow-up data were collected at four time points: (1) end of third grade (spring 2005), (2) beginning of fourth grade (fall 2005), (3) end of fourth grade (spring 2006), and (4) end of fifth grade (spring 2007). Some students stayed at the schools in the study for the full 3 years (stayers), others left (leavers) and were not followed, and new students entered the schools in the first (after the initial data collection), second, or third years of the study (new entrants). All students in each grade of the cohort were included in the sample. Table B describes the student sample overall and by program for all students and for the treatment and control groups. The study began with about 6,600 students in third grade and ended with about 6,200 in fifth grade. By the end of the fifth grade, 31 percent of the original sample had left and new entrants made up 28 percent of the fifth-graders. There were no statistically significant differences in the percentages of leavers and new entrants between the treatment and control groups overall, though there were some such differences within the individual programs.[7]

---

[4] New York University recruited 18 schools but only 14 were included in this evaluation.

[5] A longer version of table A with more detail on the programs appears as table 1 in chapter 1. Chapters 2 through 8 provide greater discussion of each program.

[6] Research teams carried out program-specific evaluations using their own procedures and measures. These are to be published separately in the literature.

[7] Greater detail on the experimental design and on the sample is provided in chapter 1 under Study Design and Methodology.

Table A.   Research teams, SACD programs, and number of schools

Research team	Program	Program features	Number of Treatment schools	Number of Control schools
University at Buffalo, State University of New York	*Academic and Behavioral Competencies Program*	Social skills training and behavior management	6	6
University of North Carolina at Chapel Hill	*Competence Support Program*	Social and emotional learning, social dynamics training, and behavior management: social information processing, social problem solving, peer networks	5	5
Vanderbilt University	*Love In a Big World*	Character education: courage, honesty, kindness, caring	6	6 [1]
Oregon State University	*Positive Action*	Social and emotional learning: values, empathy, self-control, social skills, social bonding, self-efficacy, honesty, goal setting	7	7
The Children's Institute	*Promoting Alternative Thinking Strategies*	Social and emotional learning: emotional literacy, self-control, social competence, peer relations, interpersonal problem solving	5	5
New York University	*The 4Rs Program (Reading, Writing, Respect, and Resolution)*	Conflict resolution and literacy: social problem solving, anger management, mediation	7	7
University of Maryland, College Park	*Second Step*	Violence prevention and social and emotional learning: empathy, anger management, impulse control, and problem solving	6	6

[1] Dropped to five after one control school became a magnet school and dropped out of the study prior to Year 2.
SOURCE: The Social and Character Development (SACD) Research Program.

Table B. **Student sample, overall and by program, for all students and for the treatment and control groups**

Intervention program	Year 1 (Fall 3rd grade) All students	Year 1 (Spring 3rd grade) All students	Leavers #	%[1]	New entrants #	%[2]	Year 2 (Spring 4th grade) All students	Leavers #	%[1]	New entrants #	%[2]	Year 3 (Spring 5th grade) All students	Leavers #	%[1]	New entrants #	%[2]
All programs	6,567	6,597	364	6	394	6	6,415	1,457	22	1,305	20	6,249	2,067	31	1,749	28
Treatment group	3,367	3,388	179	5	200	6	3,327	742	22	702	21	3,172	1,078	32	883	28
Control group	3,200	3,209	185	6	194	6	3,088	715	22	603	20	3,077	989	31	866	28
ABC	879	875	43	5	39	4	877	160	18	158	18	871	289	33	281	32
Treatment group	380	373	17	4	10	3**	367	72	19	59	16	353	135	36	108	31
Control group	499	502	26	5	29	6	510	88	18	99	19	518	154	31	173	33
CSP	959	975	36	4	52	5	969	230	24	240	25	947	238	25	226	24
Treatment group	476	485	20	4	29	6	474	135	28**	133	28*	458	139	29**	121	26
Control group	483	490	16	3	23	5	495	95	20	107	22	489	99	20	105	21
LBW	986	1,007	60	6	81	8	959	228	23	201	21	944	308	31	266	28
Treatment group	548	565	25	5**	42	7	556	110	20	118	21	567	145	26**	164	29
Control group	438	442	35	8	39	9	403	118	27	83	21	377	163	37	102	27
PA	811	812	74	9	75	9	764	251	31	204	27	655	408	50	252	38
Treatment group	410	416	33	8	39	9	425	108	26**	123	29	327	209	51	126	39
Control group	401	396	41	10	36	9	339	143	36	81	24	328	199	50	126	38
PATHS	786	783	39	5	36	5	778	150	19	142	18	778	243	31	235	30
Treatment group	377	374	21	6	18	5	373	66	18	62	17	378	114	30	115	30
Control group	409	409	18	4	18	4	405	84	21	80	20	400	129	32	120	30
4Rs	1,202	1,194	86	7	78	7	1,109	320	27	227	20	1,065	492	41	355	33
Treatment group	652	647	49	8	44	7	599	183	28	130	22	556	279	43	183	33
Control group	550	547	37	7	34	6	510	137	25	97	19	509	213	39	172	34

See notes at end of table.

Table B.   Student sample, overall and by program, for all students and for the treatment and control groups—Continued

Intervention program	Year 1 (Fall 3rd grade) All students	Year 1 (Spring 3rd grade) All students	Leavers #	Leavers %[1]	New entrants #	New entrants %[2]	Year 2 (Spring 4th grade) All students	Leavers #	Leavers %[1]	New entrants #	New entrants %[2]	Year 3 (Spring 5th grade) All students	Leavers #	Leavers %[1]	New entrants #	New entrants %[2]
SS	944	951	26	3	33	3	959	118	13	133	14	989	89	9	134	14
Treatment group	524	528	14	3	18	3	533	68	13	77	14	533	57	11*	66	12
Control group	420	423	12	3	15	4	426	50	12	56	13	456	32	8	68	15

* Treatment group significantly different from control group at the .05 level.

** Treatment group significantly different from control group at the .01 level.

[1] Leavers as a percentage of fall 2004 enrollment (these values are cumulative over the years).

[2] New entrants as a percentage of spring enrollment.

NOTE: Abbreviations are

    ABC: *Academic and Behavioral Competencies Program*

    CSP: *Competence Support Program*

    LBW: *Love In a Big World*

    PA: *Positive Action*

    PATHS: *Promoting Alternative Thinking Strategies*

    4Rs: *The 4Rs Program (Reading, Writing, Respect, and Resolution)*

    SS: *Second Step*

SOURCE: The Social and Character Development (SACD) Research Program.

Twenty student and school outcomes related to social and character development were used to evaluate the impact of the SACD programs on student outcomes and perceptions of school climate.[8] These were grouped into four domains: (1) Social and Emotional Competence, which contained three outcomes; (2) Behavior, which contained nine outcomes; (3) Academics, which contained two outcomes;[9] and (4) Perceptions of School Climate, which contained six outcomes. Four major data collection instruments were used to collect the scales on which the outcome measures were based. For the Child Report, students individually completed a set of 10 scales during school, and these contributed to 10 outcomes that fell across all four domains. For the Primary Caregiver Report, primary caregivers (usually parents) filled in a written survey (or completed a phone interview if they failed to complete the written version) that contained 12 scales. Six of these contributed to three outcomes in the Behavior domain (the other 6 were used to develop measures of characteristics associated with greater prevalence of child behavior problems). In the Teacher Report on Student, each student's teacher reported on 10 scales regarding a student's typical behavior in the past 30 days, and these were converted into five outcomes in the Behavior and Academics domains. In the Teacher Report on Classroom and School, the third-, fourth-, and fifth-grade teachers in a school reported on two scales that were converted into two outcomes that described their Perceptions of School Climate. In addition, they completed items that described the level of SACD activity in the classroom and school. Table C lists the 20 outcomes by domain and the reports from which they were obtained. Three of the outcomes in the Behavior domain were measured in more than one report. Altruistic Behavior and Problem Behavior were separately measured three times using responses from children, primary caregivers, and teachers; Positive Social Behavior was separately measured twice using responses from primary caregivers and teachers.

---

[8] The original scales and the outcomes are described in chapter 1 under Measures.

[9] The Academics domain was planned to but did not include student grades and standardized test scores; not all research teams were able to collect these data from their schools and districts, and the data that were collected varied in quality.

Table C.   Outcomes, by domain and data instrument

Domain/Outcome	Child Report	Primary Caregiver Report	Teacher Report on Student	Teacher Report on Classroom and School
Social and Emotional Competence Domain				
Self-Efficacy for Peer Interaction (+)	✓			
Normative Beliefs About Aggression (-)	✓			
Empathy (+)	✓			
Behavior Domain				
Altruistic Behavior (+)	✓	✓	✓	
Positive Social Behavior (+)		✓	✓	
Problem Behavior (-)	✓	✓	✓	
ADHD-Related Behavior (-)			✓	
Academics Domain				
Engagement with Learning (+)	✓			
Academic Competence and Motivation (+)			✓	
Perceptions of School Climate Domain				
Positive School Orientation (+)	✓			
Negative School Orientation (-)	✓			
Student Afraid at School (-)	✓			
Victimization at School (-)	✓			
Feelings of Safety (+)				✓
Student Support for Teachers (+)				✓

NOTE: Abbreviations are
  ADHD: Attention deficit hyperactivity disorder
  ✓: Outcome addressed
  Blank cell: Outcome not addressed
The +/- signs in parentheses indicate the direction of a beneficial outcome.
SOURCE: The Social and Character Development (SACD) Research Program.

Research Questions

The purpose of the SACD evaluation was to determine whether seven coherent, universal, school-based programs improved student social and emotional competence; improved behavior, including reducing negative behavior; improved student achievement; and improved student and teacher perceptions of school climate. The evaluation considered the programs together and individually. In addition, the evaluation considered the programs' impacts on students with different backgrounds that have been found to increase the risk of poor outcomes and possibly change student responses to the SACD programs. Also, the evaluation took into account findings from previous work showing that the level of implementation of a program affects its impact. In addition, the expectation of a positive impact on student outcomes raised the issue of whether and in what ways the programs increased the prevalence of SACD instruction. These issues led to the development of five research questions:

1. What is the average effect of the seven universal, school-based, social and character development programs on social and character development instruction in the schools?

2. What is the average effect of the seven universal, school-based, social and character development programs on students' social and emotional competence, behavior, and academics, and on perceptions of school climate?

3. What is the average effect of each specific social and character development program on students' social and emotional competence, behavior, and academics, and on perceptions of school climate?

4. Do the average effects of the seven universal, school-based social and character development programs differ by (a) students' gender and (b) students' initial risk factors (socioeconomic, family, community, and earlier child behavior)?

5. In the treatment schools, is there an association between the level of implementation of the social and character development programs and student outcomes?

Data Collection

Three issues regarding the conduct of the data collection have implications for the analysis: (1) timing of data collection, (2) percentages of the sample for which data were available for analysis, and (3) students who left the study (leavers).

Ideally, the first fall data collection would have started at the very beginning of the year to reduce the possibility that program implementation could have affected responses to the student, primary caregiver, and teacher surveys. For practical reasons, fall data collection was often delayed for several weeks to allow school populations to settle, to obtain primary caregiver consent, and to avoid disrupting planned school activities. As a result, program implementation began before initial data collection for six of the research teams (Vanderbilt University was the exception). This interval ranged from 2 to 6 weeks. In addition, at all schools teachers and principals received training on the intervention before the fall 2004 data collection. As a result, the fall 2004 reports from teachers and principals, and possibly students, are unlikely to reflect the true pre-intervention condition but instead capture what was being done at the beginning of the evaluation.[10]

Data were not successfully collected from all students, primary caregivers, and teachers. Data were not collected when written consent was not provided by primary caregivers or teachers, or when respondents refused to take part (even after consent had been given) or were unavailable at the time of data collection. Table D presents the overall consent and completion rates for each report by year and by treatment versus control group. This table also presents the percentages of the sample for which there are data for each report. These are calculated by multiplying the consent rate by the completion rate. Table D shows that 60 percent to

---

[10] Additional details on timing issues can be found under Data Collection in chapter 1.

65 percent of students had data supplied by themselves over the 3 years, 46 percent to 59 percent had data provided by their primary caregivers, and 61 percent to 67 percent had data provided by teachers. In Year 1 (third grade), a statistically significant larger percentage of the treatment group had data than the control group; there were no significant differences in Years 2 and 3. Table D also shows that data on classrooms and schools were obtained from 86 percent to 90 percent of the teachers.[11]

The evaluation did not follow all students originally assigned to the treatment or control groups (what is known as an "intent to treat" study) because data were not collected from students who left the schools. If the SACD programs caused differential student mobility in the treatment schools versus the control schools, then the impact of this mobility would be combined with the impact of the programs and the two could not be disentangled. Descriptive analyses, presented in chapter 1, did not identify statistically significant differential mobility in the treatment and control schools, but this is not definitive evidence that it does not exist.

---

[11] For additional comparisons see Consent Rates in chapter 1 and chapters 2 through 8 regarding each program.

Table D.   Consent rates, completion rates, and percentage of sample with data from each report

Report	Year 1 (Fall 3rd grade)			Year 1 (Spring 3rd grade)			Year 2 (Spring 4th grade)			Year 3 (Spring 5th grade)		
	Total	Treatment	Control	Total	Treatment	Control	Total	Treatment	Control	Total	Treatment	Control
Student sample size	6,567	3,367	3,200	6,597	3,388	3,209	6,415	3,327	3,088	6,249	3,172	3,077
Child Report (percent)												
Primary caregiver consent rate	65	67**	63	66	68**	64	67	67	66	66	67	66
Student completion rate	94	93*	94	96	96	96	95	96	95	96	97	96
Students with data[1]	61	62*	60	63	65**	61	63	65	62	64	65	63
Primary Caregiver Report (percent)												
Primary caregiver consent rate	63	64**	61	64	66**	62	64	65	63	64	65	64
Primary caregiver completion rate	92	92	92	80	80	81	78	78	77	72	71	72
Primary caregivers with data[1]	57	59*	56	51	52	50	50	51	49	46	46	46
Teacher Report on Student (percent)												
Primary caregiver consent rate[2]	65	67**	63	66	68**	64	67	67	66	66	67	66
Teacher completion rate	96	96	96	99	99	99	100	100**	99	98	98	99
Students with data[1]	62	64**	61	65	67**	63	66	67	65	65	66	65
Teacher Report on Classroom and School (3rd- to 5th-grade teachers) (percent)												
Teacher consent rate	96	98***	92	95	97*	94	95	97	94	96	97	95
Teacher completion rate	91	90	93	91	90	91	94	94	94	92	91	93
Teachers with data[1]	87	88	86	87	88	86	90	90	89	89	88	89

* Treatment group significantly different than control group at the .05 level.
** Treatment group significantly different than control group at the .01 level.
*** Treatment group significantly different than control group at the .001 level.
[1] Calculated as consent rate x completion rate.
[2] The primary caregiver consent rates for the Child Report and the Teacher Report on Student are identical, as the primary caregiver gave consent to both together.
SOURCE: The Social and Character Development (SACD) Research Program.

Initial Characteristics

An examination of the initial characteristics of the students, families, teachers, and schools found that the treatment and control groups were similar on a set of observed characteristics (with the exception of the use of SACD activities in the schools), providing evidence that the random assignment of schools within programs created similar groups.[12] The data for this examination were collected in fall 2004 from enrolled third-grade students, their primary caregivers, and their third-grade teachers. In addition, third-, fourth-, and fifth-grade teachers and principals in the study schools provided information about SACD activities being used in the classroom and school. The initial data were collected after the staff at the treatment schools began receiving training in their programs and, for six of the programs, after implementation had begun.

The sample's treatment and control groups were similar along the observed student, primary caregiver, and community characteristics. These included (1) student gender, race/ethnicity, and age; (2) primary caregiver race/ethnicity, age, employment, marital status, education, and household income; and (3) community risks and resources. In addition, there were no significant differences between the treatment and control groups for the 20 outcome measures and the five measures of initial risk, showing that the two groups of students started, on average, at the same place in third grade.

The teachers in the treatment and control schools were similar in gender, race/ethnicity, years of teaching, and certification. The only statistically significant difference concerned degree attainment, as a larger percentage of treatment teachers (60%) had a master's or doctoral degree than did control teachers (52%). There was no significant difference between the treatment and control schools with regard to student composition (race/ethnicity and school lunch eligibility), number of students, number of full-time teachers, Title I status, highest and lowest grades, urbanicity, and number of years the principal had been at the school. There was also no significant difference in teacher reports on nine dimensions concerning their school environment: feelings of safety, resource adequacy, student support, staff freedom to teach as desired, affiliation with and ties to colleagues, innovation regarding trying new approaches to teaching, professional interest, participatory decisionmaking, and work pressure.

An examination of the initial level of activities to promote social and character development in the classroom and schoolwide, and of the materials and methods used in these activities, revealed that the control condition for the evaluation was not a "no treatment" control. Rather, it was a "standard practice" control condition, in which more than half of teachers and 80 percent to 90 percent of principals reported schoolwide and classroom activities designed to promote social and character development. Standard practice at the control schools included the use of specific materials and practices to promote social and character development, as well as professional development related to social and character development of students for staff. Many kinds of activities and strategies were provided at rates and in types and amounts similar to those reported in treatment schools.

Treatment teachers reported greater use of and training in SACD activities than control teachers more often than would be expected by chance. This may reflect actual differences in the use of SACD activities prior to implementation of the programs, or it may reflect that program implementation (for six programs) and program training for staff had started before initial data collection.[13]

---

[12] For greater detail, see Initial Characteristics in chapter 1.

[13] For details on these comparisons, see The Initial Level of Social and Character Development in the Schools in chapter 1.

Analysis and Results

Four sets of analyses were done to evaluate the impacts of the SACD programs.[14] First, the effects of the adoption of the programs on the use of SACD activities in the classroom and school were examined on an annual basis for Years 1 to 3. Second, the programs' impacts on student outcomes and perceptions of school climate were analyzed at the end of each of the 3 years of the study (i.e., from the start of third grade to the end of third grade, to the end of fourth grade, and to the end of fifth grade). Third, subgroup analyses were done on the same annual basis to examine whether the programs' impacts on student outcomes, when combined, differed by four subgroups: (1) girls versus boys, (2) stayers versus new entrants, (3) students with different levels of baseline risk, and (4) students in treatment schools with high versus low fidelity of implementation. Fourth, a growth curve analysis was done to determine impacts on annual student growth in the outcomes over the 3 years. All but the third analysis were done for the combination of all seven programs and for each program separately.

Annual Impacts on the Use of Social and Character Development Activities

The SACD programs were expected to increase the use of activities to promote social and character development in the classroom and school. The analysis of activity use was based on data from the Teacher Report on Classroom and School (TRCS), which was filled out by all the third-, fourth-, and fifth-grade teachers. Data from the TRCS were used to create 83 SACD activity outcome measures, and these were grouped into six domains: (1) teacher use of any activities to promote social and character development, which had 16 variables; (2) teacher use of any activities associated with a specific SACD program (known as a "named" program), which had 14 variables; (3) teacher use of materials and classroom strategies for SACD activities, which had 29 variables; (4) use of schoolwide strategies, which had 6 variables; (5) teacher involvement in related professional development, which had 9 variables; and (6) teacher support for SACD efforts in the school and the prevalence of an environment conducive to the social and character development of students, which had 9 variables.

To estimate the impact of the SACD programs on use of SACD activities, the statistical significance of the differences in means between the treatment and control groups was tested for each of the 83 SACD activity outcome measures.[15] For the overall analysis, the data were first weighted to give equal weight to each program and to each school within a site. For the program-by-program analyses, each school within a site received equal weight (school weights differed between program analyses, as the number of schools was not constant among programs).

In addition to estimating the impacts of the SACD programs on the individual outcome measures, the impacts on the six domains were also examined. Testing the impact on the domains was done to adjust for the multiple comparisons made within each domain in order to address the increased chances of finding a spurious outcome when more than one test was done. As a result, two sets of results were obtained: (1) the

---

[14] The analytical techniques used for each of the four sets of analyses are summarized in the following sections and discussed in detail in chapter 1.

[15] Three factors contributed to the decision to use differences in means. First, because of random assignment, simple treatment-control contrasts provided unbiased estimates of program impacts. Second, only initial values (rather than true baseline values) for these outcomes were available to use in a model because training (at all treatment schools) and program implementation (at 36 treatment schools) began before data collection. The decision to use initial values in an analysis partly depends on whether the initial training and implementation occurring before data collection would be expected to have immediate and large impacts on the outcomes (Schochet 2008b). For this analysis, the outcomes are based on teacher actions and so would likely be upwardly influenced by the teacher training and short period of teacher implementation before pretesting (in contrast to student outcomes, which would be less likely to be so influenced). For this reason, a model-based analysis using the initial values as covariates was not chosen. Third, preliminary analyses indicated no gain in precision from the inclusion of other covariates.

impacts on the individual outcomes unadjusted for multiple comparisons, and (2) the impacts on the domains that serve as the multiple comparison adjustment. To test whether the SACD programs had a statistically significant impact on each of the six domains, a set of three heuristics was used and a significant effect on the domain was found if any one of the heuristics was met.[16]

The results provide evidence that the SACD programs increased the reported implementation of SACD activities in the classroom. Over the 3 years, 249 comparisons (83 outcome variables times 3 years) of treatment and control teachers were tested, with 12 to 13 expected to be statistically significant by chance. The analysis found 127 comparisons were statistically significant, with all showing greater reported use of SACD activities by treatment teachers. When examining the individual outcomes within each domain, the analysis found the following: (1) treatment teachers reported significantly greater use of SACD activities in the classroom for 31 of the 48 comparisons, (2) treatment teachers reported significantly greater use of SACD activities linked to a named SACD program for 39 of the 42 comparisons, (3) treatment teachers reported significantly greater use of materials and instructional methods to promote social and character development for 40 of the 87 comparisons, (4) there were no significant differences for the 18 comparisons made regarding the use of schoolwide strategies, (5) treatment teachers reported significantly greater receipt of training to promote social and character development for 13 of the 27 comparisons, and (6) treatment teachers reported significantly greater use of practices conducive to social and character development but similar attitudes toward it for 4 of the 27 comparisons.[17]

The results from the analysis of the individual outcomes also provide further evidence that the control group was a "standard practice" rather than a "no treatment" control. Over the 3 years, control teachers continued to report use of SACD activities. For example, over the 3 years, 86 percent to 90 percent of control teachers reported using a SACD activity to address any one of the six SACD goals versus 95 percent to 96 percent of the treatment teachers. Similarly, over the 3 years, 20 percent to 36 percent of control teachers reported using a SACD activity linked to a named SACD program to address any one of the six SACD goals versus 68 percent to 72 percent of the treatment teachers.[18]

The findings from the analysis of the six domains are consistent with the results from the individual outcome analysis. For all 3 years, treatment teachers reported statistically significant greater implementation in four of the six SACD activity domains: (1) use of any SACD activities in the classroom, (2) use of SACD activities from named programs in the classroom, (3) use of materials and teaching strategies for SACD activities, and (4) participation in relevant professional development. No evidence was found that the programs affected the other two domains—the use of schoolwide strategies and attitudes and practices that create an environment conducive to students' social and character development. Table E shows where statistically significant impacts on the use of SACD activities occurred by domain for all seven programs and for each program by year. In table E, a plus sign indicates a significant positive impact on the domain, and superscript numerals show which heuristics identified the domain as significant.

---

[16] The three heuristics included (1) determining if more than half of the individual outcomes within a domain had a similar and statistically significant impact, (2) doing one overall test of impact on all outcomes within a domain, and (3) checking for any statistically significant results among the outcomes within a domain after applying a specific statistical adjustment for multiple comparisons (Benjamini-Hochberg 1995) to each one. See Year-by-Year Impacts on Use of Social and Character Development Activities in chapter 1 for longer descriptions of these heuristics and how they were used.

[17] For more details on the impacts on the individual outcomes making up the domains see Year-by-Year Impacts on Use of Social and Character Development Activities in chapter 1.

[18] For more details on the percent of treatment and control teachers reporting on the use of SACD activities see Year-by-Year Impacts on Use of Social and Character Development Activities in chapter 1.

Table E.   Significant impacts on use of SACD activity domains, overall and by program

Program	SACD activities	SACD activities linked to named programs	Classroom materials and strategies	Schoolwide strategies	Professional development	Attitudes and practices
Overall						
Year 1	+[1,2,3]	+[1,2,3]	+[2,3]		+[1,2,3]	
Year 2	+[1,2,3]	+[1,2,3]	+[1,2,3]		+[2,3]	
Year 3	+[1,2,3]	+[1,2,3]	+[1,2,3]		+[2,3]	
ABC						
Year 1	+[2]			+[1]		
Year 2						
Year 3						+[3]
CSP						
Year 1	+[2]	+[1,3]			+[3]	
Year 2		+[3]				
Year 3		+[3]				
LBW						
Year 1	+[2,3]	+[3]	+[3]		+[3]	
Year 2	+[3]	+[3]		+[2]		
Year 3						
PA						
Year 1	+[2,3]	+[1,3]	+[3]			+[3]
Year 2		+[1,3]				
Year 3	+[3]	+[3]				
PATHS						
Year 1	+[2,3]	+[1,3]			+[1,3]	
Year 2	+[3]	+[1,3]				
Year 3		+[1,3]				
4Rs						
Year 1	+[1,2,3]	+[1,2,3]	+[3]		+[1,3]	
Year 2	+[1,3]	+[1,3]				
Year 3	+[1,3]	+[1,2,3]	+[3]		+[3]	

See notes at end of table.

Table E. **Significant impacts on use of SACD activity domains, overall and by program—Continued**

| Program | SACD activity domain | | | | | |
	SACD activities	SACD activities linked to named programs	Classroom materials and strategies	Schoolwide strategies	Professional development	Attitudes and practices
SS						
Year 1	+[3]	+[1,3]	+[3]			
Year 2	+[1,3]	+[1,3]	+[3]			
Year 3	+[3]	+[1,3]	+[3]			

[1] Based on univariate statistical tests, at least half of the impacts were positive and statistically significant and no impact was negative and statistically significant.
[2] The omnibus impact for all the outcomes measured together was positive and statistically significant on the basis of a multivariate statistical test.
[3] At least one outcome remained positive and statistically significant and no outcome was negative and statistically significant after applying the Benjamini-Hochberg (1995) procedure to adjust significance levels downward to account for the multiple testing of impacts.
NOTE: Abbreviations are
　ABC: *Academic and Behavioral Competencies Program*
　CSP: *Competence Support Program*
　LBW: *Love In a Big World*
　PA: *Positive Action*
　PATHS: *Promoting Alternative Thinking Strategies*
　4Rs: *The 4Rs Program (Reading, Writing, Respect, and Resolution)*
　SS: *Second Step*
Abbreviations of the findings are
　+: Statistically significant beneficial impact on domain
　Blank cell: Finding of no statistically significant impact
Significance is based on $p \leq .05$. No detrimental impact was found statistically significant at or below the .05 level. Description of SACD Activity Domains and the heuristics used to determine the statistically significant beneficial impact on the domain (for more detail, see the Measures section in chapter 1):
　SACD activities: based on 16 teacher-reported measures on the use of SACD activities in the classroom.
　SACD activities linked to named programs: based on 14 teacher-reported measures on the use of SACD activities associated with a named program in the classroom.
　Classroom materials and strategies: based on 29 teacher-reported measures, 7 concerning materials used in the classroom and 22 concerning classroom strategies.
　Schoolwide strategies: based on six teacher-reported measures concerning strategies to promote SACD schoolwide.
　Professional development: based on nine teacher-reported measures concerning their participation in SACD-related training.
　Attitudes and practices: based on nine teacher-reported measures, three concerning teacher attitudes toward SACD efforts in the school and six concerning school practices conducive to the social and character development of students.
SOURCE: The Social and Character Development (SACD) Research Program.

Annual Impacts on Student Outcomes and Perceptions of School Climate

The SACD programs were expected to improve children's social and emotional competence, behavior, academics, and perceptions of school climate as measured by 20 outcome variables. One test of these hypotheses was to examine the year-by-year impacts of the SACD programs on these outcomes over the 3 years as the students progressed from third through fifth grades. The examination of year-by-year impacts entailed three sets of analyses resulting in three sets of impacts. The first set of analyses compared the outcomes of treatment and control students from the fall of third grade to the spring of third grade. The second set compared the outcomes from the fall of third grade to the spring of fourth grade, and the third set compared outcomes from the fall of third grade to the spring of fifth grade. Within each set of year-by-year analyses, an analysis of all the programs together provided impact results for the set of seven coherent, universal, school-based programs, and separate analyses of each individual program provided results specific to each program. The combined analysis was able to detect smaller statistically significant impacts, because of

its larger sample size and the associated greater power, than the analyses of each SACD program, which were based on smaller samples.[19]

The random assignment of schools ensured that unbiased estimates of the average impacts of the SACD programs (relative to the social and character development activities offered in the control schools) could be computed as the differences in the average outcomes of students and teachers in the treatment and control schools. However, regression procedures were used rather than simple differences-in-means procedures to estimate impacts to improve the statistical precision of the estimates; to address the clustering of students within schools; and to adjust for differences between treatment and control group observable characteristics due to random selection, study nonconsent, and interview nonresponse. A hierarchical linear model (HLM) was used to estimate regression-adjusted impacts (Bryk and Raudenbush 1992). The basic model consisted of two levels that were indexed by students or teachers and by schools. The model included covariates that adjusted for statistically significant treatment and control differences at initial data collection. These covariates were chosen because they had predictive power across a broad range of outcomes. Sample weights were used in the analyses in order to (1) give each *site* equal weight in the calculation of pooled impact estimates, (2) give each *school* equal weight in each site, and (3) adjust for missing outcome data due to nonconsent and nonresponse. The model was estimated using data from all seven programs together (using all three types of weights) and individually for each program (using the second and third types of weights). A set of analyses to examine differences by subgroup for the combined data was done by examining the significance of a coefficient on an interaction term between the treatment status indicator variable and subgroup indicator variable (or multiple coefficients and multiple interaction terms when there were more than two subgroups) when the subgroup variables were included. The association of fidelity of implementation with the outcomes was examined in a similar way, using a fidelity indicator variable and an interaction term between the fidelity variable and the treatment status variable.[20]

Results from estimating the model were provided in effect sizes.[21] A standard two-tailed test was used to determine the *p*-value for the coefficient of each outcome measure. Coefficients with *p*-values of .05 or below were considered statistically significant and identified as such. Impacts that were not statistically significant but were .25 standard deviation units or more in magnitude were identified as "substantively important," following the practice used by the What Works Clearinghouse.[22] Substantively important impacts identify effects that may be large enough to have practical importance but are not found to be statistically significant, potentially because of sample size constraints.

The 20 outcome variables were grouped under a set of four domains: (1) Social and Emotional Competence, which contained 3 outcomes; (2) Behavior, which contained 9 outcomes; (3) Academics, which contained 2 outcomes; and (4) Perceptions of School Climate, which contained 6 outcomes. As in the case of the SACD activities domains, a set of heuristics was used to test the significance of the impacts on the four outcome

---

[19] The combined analysis provided a sample size sufficient to detect student-level impacts at minimum detectable effect sizes ranging from 0.03 to 0.24 standard deviations (see table 1.25 in chapter 1 for details).

[20] For a discussion of the regression model, see Analysis under the section Year-by-Year Impacts on Use of Social and Character Development Activities in chapter 1.

[21] Effect sizes were calculated by dividing the estimated impact (the coefficient estimated by the regression model) by the standard deviation of the outcome measure for the control group. The standard deviation was calculated using data for the weighted control group. It was calculated at the time of data collection for which the effect size (impact) was estimated.

[22] The What Works Clearinghouse was established in 2002 by the U.S. Department of Education's Institute of Education Sciences to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence of what works in education.

domains to adjust for the multiple comparisons made within each domain.[23] The analysis of the year-by-year impacts then produced two sets of results: (1) the impacts on the individual outcomes unadjusted for multiple comparisons, and (2) the impacts on the domains that served as the multiple comparison adjustment.

Column 2 of table F reports the number of statistically significant and substantively important impacts found for the year-by-year analysis of all seven programs together. Specifically, 2 of 60 estimated impacts on the outcomes for the seven SACD programs combined for the 3 years were found statistically significant (versus 3 expected by chance). The 2 statistically significant impacts were that the combined SACD programs had a beneficial impact on the teacher-reported measure for Student Support for Teachers in Years 1 and 2, with effect sizes of 0.12 and 0.16, respectively. None of the remaining 58 estimated impacts were found to be substantively important.[24] In addition, 12 impacts on the domains were estimated (4 domains times 3 years), with 1 expected to be statistically significant by chance. The analysis found 2 significant negative impacts on the domain of Social and Emotional Competence in Years 2 and 3 (these data are not shown in a table). The results for the individual outcomes and the domains provide no evidence that the SACD programs improved student outcomes and perceptions of school climate.

The lack of statistically significant impacts found in the combined-program analysis was not due to offsetting beneficial and detrimental impacts among the individual programs. For the individual SACD programs, 16 significant impacts were found over the 3 years (9 beneficial and 7 detrimental) versus 21 expected by chance from the 420 statistical tests done (see column 3 of table F). In only one program did a significant impact occur on an outcome in more than one year, and in this case the impact was beneficial in Year 1 and detrimental in Year 2. In addition, 19 substantively important impacts were found (10 beneficial and 9 detrimental). There was little replication of the substantively important impacts: one program had a substantively important beneficial impact on one outcome for all 3 years. Table G identifies the statistically significant and nonsignificant substantively important results by program, outcome, and year. It provides a visual view of the balance between beneficial and detrimental impacts and of the lack of persistence in impacts by program across the years.

---

[23] The heuristics were used to determine whether the results for the multiple outcomes within each domain showed a statistically significant impact on the domain as a whole. A significant effect on the domain was found if any one of the heuristics was met. In addition to the three heuristics mentioned in footnote 16, a fourth heuristic (not applicable to the analysis of teacher-reported SACD activities) was used. For this heuristic, the statistical model used to estimate impacts on the individual outcomes was re-estimated using a composite of all the outcome variables within a domain. The composite was formed by standardizing each outcome variable using its standard deviation, combining the values of the outcome variables, and taking the average of the final value.

[24] For the actual effect size and significance level of each outcome, see table 1.26 in chapter 1.

Table F.   Significant impacts on student outcomes, overall and by program, for all 3 years

Domain (number of outcomes)	Combined-program analysis (all seven programs together) (60 impact estimates)[1]	Individual program analyses (each program separately) (420 impact estimates)[2]
Total		
Statistically significant outcomes	2 (Beneficial)	16 (9 Beneficial, 7 Detrimental)
Substantively important outcomes	0	19 (10 Beneficial, 9 Detrimental)
Social and Emotional Competence (3)		
Statistically significant outcomes	0	0
Substantively important outcomes	0	0
Behavior (9)		
Statistically significant outcomes	0	8 (6 Beneficial, 2 Detrimental)
Substantively important outcomes	0	6 (2 Beneficial, 4 Detrimental)
Academics (2)		
Statistically significant outcomes	0	4 (1 Beneficial, 3 Detrimental)
Substantively important outcomes	0	0
Perceptions of School Climate (6)		
Statistically significant outcomes	2 (Beneficial)	4 (2 Beneficial, 2 Detrimental)
Substantively important outcomes	0	13 (8 Beneficial, 5 Detrimental)

[1] For each year, 20 impacts (on 20 outcomes) were estimated and 1 statistically significant impact would be expected each year by chance (for a total of 3).

[2] For each year, 140 impacts were estimated (7 programs by 20 outcomes) and 7 statistically significant impacts would be expected each year by chance (for a total of 21).

NOTE: For each outcome, a finding of "beneficial" indicates the program(s) had a beneficial impact on that particular outcome; a finding of "detrimental" indicates a detrimental impact on that outcome. Significance is based on $p \leq .05$. The number of results found significant was no more than expected by chance.

SOURCE: The Social and Character Development (SACD) Research Program.

Table G. Individual program statistically significant impacts and nonsignificant but substantively important impacts

| | Statistically significant[1] | | Nonstatistically significant but substantive[2] | |
| | Beneficial impacts | Detrimental impacts | Beneficial impacts | Detrimental impacts |
Program	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)
Total	9	7	10	9
Year 1	3	2	2	0
Year 2	5	2	6	1
Year 3	1	3	2	8
ABC				
Year 1	Altruistic Behavior (TRS) (.39) (.026)			
Year 2	Academic Competence (CR) (.31) (.011) Feelings of Safety (TRCS) (.75) (.003)	Altruistic Behavior (CR) (-.20) (0.029)	Student Support for Teachers (TRCS) (.27) (.276)	
Year 3	Positive Social Behavior (PCR) (.21) (.041)		Feelings of Safety (TRCS) (.31) (.235)	
CSP				
Year 1				
Year 2	Problem Behavior (PCR) (-.21) (.042)		Altruistic Behavior (TRS) (.47) (.132) Student Afraid at School (CR) (-.26) (.090)	
Year 3				Altruistic Behavior (TRS) (-.41) (.132) Feelings of Safety (TRCS) (-.36) (.246)
LBW				
Year 1	Altruistic Behavior (PCR) (.31) (.005) Student Support for Teachers (TRCS) (.52) (.022)			
Year 2			Student Support for Teachers (TRCS) (.28) (.428)	Altruistic Behavior (TRS) (-.34) (.270)

See notes at end of table.

| | Statistically significant[1] | | Nonstatistically significant but substantive[2] | |
| | Beneficial impacts | Detrimental impacts | Beneficial impacts | Detrimental impacts |
Program	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)
Year 3		Engagement with Learning (CR) (-.35) (.030) Positive School Orientation (CR) (-.33) (.047) Feelings of Safety (TRCS) (-.70) (.046)		Problem Behavior (CR) (.31) (.223) Student Support for Teachers (TRCS) (-.26) (.543)
PA				
Year 1		Engagement with Learning (CR) (-.25) (.017)	Altruistic Behavior (TRS) (.27) (.480)	
Year 2	Positive Social Behavior (PCR) (.24) (.039) Problem Behavior (TRS) (-.24) (.048)		Student Support for Teachers (TRCS) (.28) (.113)	
Year 3				
PATHS				
Year 1				
Year 2				
Year 3				Altruistic Behavior (TRS) (-.31) (.485) Feelings of Safety (TRS) (-.29) (.582)
4Rs				
Year 1		Academic Competence (CR) (-.17) (.032)		
Year 2				
Year 3				Feelings of Safety (TRS) (-.42) (.146) Student Support for Teachers (TRCS) (-.35) (.109)

See notes at end of table.

Table G.   Individual program statistically significant impacts and nonsignificant but substantively important impacts—Continued

	Statistically significant[1]		Nonstatistically significant but substantive[2]	
	Beneficial impacts	Detrimental impacts	Beneficial impacts	Detrimental impacts
Program	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)	(Report) (Effect size) (*p*-value)
SS				
Year 1			Feelings of Safety (TRCS) (.37) (.216)	
Year 2		Positive Social Behavior (PCR) (-.14) (.050)	Feelings of Safety (TRCS) (.39) (.197)	
Year 3			Feelings of Safety (TRCS) (.52) (.062)	

[1] Out of the 140 comparisons made for each year, 7 would be expected to be statistically significant at the .05 level by chance (for a total of 21).
[2] Defined as impacts that were not statistically significant but were .25 standard deviation units (absolute value) or more in magnitude.
NOTE: Abbreviations are
  ABC: *Academic and Behavioral Competencies Program*
  CSP: *Competence Support Program*
  LBW: *Love In a Big World*
  PA: *Positive Action*
  PATHS: *Promoting Alternative Thinking Strategies*
  4Rs: *The 4Rs Program (Reading, Writing, Respect, and Resolution)*
  SS: *Second Step*
  CR: Child Report
  PCR: Primary Caregiver Report
  TRS: Teacher Report on Student
  TRCS: Teacher Report on Classroom and School
  Blank cell: Finding of no impact
All impact estimates were calculated using regression models in which each school within a program was weighted equally. The standard errors of all estimates account for design effects due to unequal weighting and the clustering of students within schools. Significance is based on *p* ≤ .05. The number of results found significant was no more than expected by chance.
SOURCE: The Social and Character Development (SACD) Research Program.

The lack of statistically significant impacts found in the combined-program analysis was not due to offsetting impacts among subgroups defined by gender, stayer versus new entrant status, and different levels of initial student risk. A greater number of significant differences were found than would be expected by chance in the gender and initial student risk analyses (but not in the stayer analysis); however, the differences did not favor any one subgroup, and only in seven cases were they replicated across the years (these data are not shown in a table). The analysis by gender found 8 statistically significant differences between the genders out of 54 possible, where 3 would be expected by chance; half of the significant impacts were beneficial impacts and half were detrimental impacts (one of the detrimental impacts was replicated in Years 2 and 3). The analysis of stayers versus new entrants found no statistically significant differences in the 36 comparisons. For the five types of different initial risk, the analysis found 41 statistically significant differences among three levels of student risk (low, average, and high) out of the 270 possible (13 or 14 would have been expected to be significant by chance); of these, 26 showed more beneficial impacts for higher risk students versus lower risk students (4 of these were replications across years), and 15 showed more detrimental impacts for higher risk students (2 of these were replications across years).[25]

The analysis of the fidelity data found little evidence of a relationship between high fidelity and more beneficial outcomes (these data are not shown in a table). The number of significant associations found between fidelity and beneficial outcomes was higher than expected by chance (5 associations found compared to 3 that might be expected by chance out of 54 estimated impacts) but 4 of the 5 significant results were due to detrimental associations between low fidelity and outcomes (rather than beneficial associations between high fidelity and outcomes).

In conclusion, the analysis of the year-by-year impacts did not yield evidence that the seven SACD programs, combined and individually, improved student social and character development. A small number of findings were statistically significant (but no more than would be expected by chance, except for several of the subgroups) or substantively important. These results were split into similar numbers of beneficial and detrimental impacts; that is, the SACD programs improved some outcomes but worsened others. In the majority of cases, the results (both beneficial and detrimental) occurred in only 1 year and were not replicated across the 3 years of the study.

Impacts on Growth of Student Outcomes

A growth curve analysis was done to examine the change over time in the impacts on the outcomes between fall 2004 and spring 2007.[26] The growth curve analysis used the same covariates and compared results across the same subgroups as the cross-sectional analyses to ensure the comparability of the results. However, it differed from the cross-sectional analyses by examining the estimated impacts on the trajectories of student outcomes over time, rather than at a point in time. The sample of students for the growth curve analysis included all students who were enrolled in one of the study schools during the study period and who completed a survey during the initial data collection or at any of the four follow-up survey points. The percentages of the sample with responses were similar for treatment and control schools in most survey waves; however, the percentages were statistically significantly higher for the treatment group on the Child Report and Primary Caregiver Report in fall 2004 and on the Child Report in spring 2005. There was considerable turnover within the sample by spring 2007. Across the three survey instruments that reported on students, about two-thirds of the sample had taken part in the original fall 2004 survey (66% for the Child Report and the Primary Caregiver Report, and 68% for the Teacher Report on Student). There were no statistically significant differences in the level of turnover by treatment status.[27] The 18 child-level outcome measures collected from the Child Report, Primary Caregiver Report, and Teacher Report on Student were used in the growth curve analysis. The 2 other outcomes, Feelings of Safety and Student Support for Teacher,

---

[25] The section Year-By-Year Subgroup Analysis of Impacts on Students in chapter 1 provides the details on these results.

[26] See Impacts on Growth of Student Outcomes in chapter 1 for more detail on the growth curve analysis and results.

[27] For details see table 1.36 in chapter 1.

collected by the Teacher Report on Classroom and School were not used because they were measured at the teacher level.

The impacts over time were estimated using growth curve models (with time since implementation of the program as the time metric) by examining treatment and control group differences in the trajectories of student outcomes during the study while accounting for clustering at the school level. The growth curve models were estimated using a three-level hierarchical linear model, where Level 1 corresponded to time since implementation of the program, Level 2 to students, and Level 3 to schools. The models included the same set of covariates that were used for the cross-sectional analysis except they excluded the initial outcome measure as a covariate because it was used as the outcome measure for the growth curve analysis at time 1 (fall 2004). Similarly, sample weights were used in all analyses to (1) give each program equal weight within each time period; (2) give each school equal weight in each program (within each time period); and (3) give each time period equal weight in the analysis. However, the weights were not adjusted for consent and response differences across classrooms or schools because the population of students within the schools changed over time as students entered and left the schools. Similar to the year-by-year analyses, the growth model was estimated for the seven programs combined and for each program individually. In addition, the effects of the combined programs by subgroup were estimated by including interaction terms between treatment status, time since implementation, and indicators of membership in subgroups.

Growth curve effect sizes were calculated by dividing the estimated impact of the treatment on the outcome growth trajectory by the standard deviation of that outcome.[28] The growth curve estimates the change in outcomes over 1 year, so the estimated impact of the treatment on the growth trajectory equals the difference between the treatment group's outcome and the control group's outcome, on average, after 1 year of the study.[29] The effect size measures the number of standard deviations the treatment group differs from the control group after 1 average year of the study, making it analogous to the effect size calculations for the cross-sectional analysis.

The growth curve analysis found no significant effects of the seven SACD programs when combined (table H). None of the 18 estimated impacts on the trajectories of child outcomes from the average of the seven SACD programs were statistically significant. The estimated effect sizes all fell below .07 (absolute value).

The lack of significant effects reflected in the analysis of the seven programs together was not found to be due to differences among the individual programs. The results from the analysis of individual programs indicate that the lack of significant impacts in the overall evaluation reflected the lack of significant impacts at the program level. Six statistically significant impacts were found in the program analyses, the same number expected by chance given that 126 impacts were estimated. Two were beneficial impacts and four were detrimental impacts (table I).

---

[28] The standard deviation was calculated using data for the weighted control group. It was calculated at the time of data collection for which the effect size (impact) was estimated.

[29] This impact estimate takes into account differences between the initial levels of the outcome for the treatment and control groups, differences in their covariates, and the effects of clustering at the school level.

Table H.   Impacts on growth of child outcomes from combined-program analysis

Scale–Report	Mean score at implementation[2]	Average growth in the score per year[1]		Impact on growth[3]	Effect size[4]	Standard error of impact	p-value of impact
		Treatment group	Control group				
Social and Emotional Competence Domain							
Self-Efficacy for Peer Interactions–CR (+)	2.95	0.13	0.13	0.00	0.00	0.01	0.942
Normative Beliefs About Aggression–CR (-)	1.23	0.10	0.07	0.02	0.04	0.01	0.115
Empathy–CR (+)	2.41	-0.14^	-0.12	-0.02	-0.05	0.01	0.070
Behavior Domain							
Altruistic Behavior–CR (+)	1.41	-0.17	-0.16	-0.01	-0.01	0.02	0.681
Altruistic Behavior–PCR (+)	2.32	-0.03	-0.03	0.01	0.01	0.01	0.616
Altruistic Behavior–TRS (+)	1.40	-0.04	0.00	-0.04	-0.07	0.03	0.224
Positive Social Behavior–PCR (+)	2.99	0.04	0.03	0.01	0.02	0.01	0.217
Positive Social Behavior–TRS (+)	3.00	0.00	0.01	-0.01	-0.01	0.03	0.671
Problem Behavior–CR (-)	0.24	0.10	0.08	0.01	0.02	0.01	0.351
Problem Behavior–PCR (-)	1.58	-0.01	-0.01	0.00	0.00	0.01	0.843
Problem Behavior–TRS (-)	1.38	0.05	0.05	0.00	0.00	0.01	0.909
ADHD-Related Behavior–TRS (-)	1.75	-0.02	-0.01	-0.01	-0.01	0.02	0.707
Academics Domain							
Engagement with Learning–CR (+)	3.69	-0.03	-0.03	0.00	-0.01	0.01	0.707
Academic Competence and Motivation–TRS (+)	2.87	0.02	0.03	-0.01	-0.01	0.02	0.590

See notes at end of table.

Table H.   Impacts on growth of child outcomes from combined-program analysis—Continued

Scale–Report	Mean score at implementation[2]	Average growth in the score per year[1]					
		Treatment group	Control group	Impact on growth[3]	Effect size[4]	Standard error of impact	*p*-value of impact
Perceptions of School Climate Domain							
Positive School Orientation–CR (+)	3.09	-0.24	-0.21	-0.03	-0.03	0.02	0.163
Negative School Orientation–CR (-)	1.84	0.10	0.09	0.01	0.01	0.02	0.696
Student Afraid at School–CR (-)	2.38	-0.08	-0.08	0.00	0.00	0.02	0.956
Victimization at School–CR (-)	0.76	-0.03	-0.02	-0.01	-0.01	0.02	0.368

^ Treatment group significantly different from control group at the .10 to > .05 level.

[1] Pertains to the estimated slope of the outcome for the treatment or control groups.

[2] The average score at implementation is calculated across treatment and control groups, using regression models for adjustment on covariates.

[3] Estimated difference between the slope of the treatment and control groups.

[4] The slope of the treatment group minus the slope of the control group divided by the standard deviation of the outcome for the combined-program control group (the standard deviation is calculated without accounting for school-level clustering or regression adjustments).

NOTE: Abbreviations are
  CR: Child Report
  PCR: Primary Caregiver Report
  TRS: Teacher Report on Student
  ADHD: Attention deficit hyperactivity disorder

The +/- signs in parentheses indicate the direction of a beneficial outcome. No findings were found statistically significant at or below the .05 level. All impact estimates were calculated using HLM 6.06.

SOURCE: The Social and Character Development (SACD) Research Program.

Table I.   **Significant impacts from the growth curve analyses of the individual programs**

Program	Significant beneficial impacts (Report) (Effect size) (*p*-value)	Significant detrimental impacts (Report) (Effect size) (*p*-value)	Total significant*[1] impacts
Total	2	4	6
ABC			0
CSP	Victimization at School (CR) (-.09) (.050)		1
LBW		Positive School Orientation (CR) (-.13) (.016)	1
PA			0
PATHS	Academic Competence (TRS) (.08) (.048)		1
4Rs			0
SS		Engagement with Learning (CR) (-.09) (.021) Positive Social Behavior (TRS) (-.19) (.019) Empathy (CR) (-0.13) (.028)	3

* Significantly different from zero at the .05 level.

[1] Out of the 126 comparisons made (7 programs times 18 outcomes), 6 would be expected to be statistically significant at the .05 level by chance.

NOTE: Abbreviations are
  ABC: *Academic and Behavioral Competencies Program*
  CSP: *Competence Support Program*
  LBW: *Love In a Big World*
  PA: *Positive Action*
  PATHS: *Promoting Alternative Thinking Strategies*
  4Rs: *The 4Rs Program (Reading, Writing, Respect, and Resolution)*
  SS: *Second Step*
  CR: Child Report
  TRS: Teacher Report on Student
  Blank cell: Finding of no impact
The number of results found significant was no more than expected by chance.
SOURCE: The Social and Character Development (SACD) Research Program.

The lack of findings from the combined-program analysis was not found to be due to differences among subgroups (these data are not shown in a table). The subgroup analyses did not find evidence regarding differential impacts of the seven SACD programs, combined, on the subgroups. For the gender subgroup analysis, none of the 36 estimated impacts were statistically significant. For the analysis of new entrants, 1 impact was found to differ significantly and detrimentally from the impact on members of the original sample, which was no more than would be expected by chance. There were 6 outcomes out of 90 estimated for which growth trajectories differed significantly across initial risk levels (4 to 5 would have been expected by chance); for 4 of the 6 outcomes, the programs had more beneficial impacts for children with higher initial risk levels, while for 2 outcomes they had more detrimental impacts for children with higher initial risk levels.

Discussion

The year-by-year analysis and the growth curve analysis did not find that the seven SACD programs improved student outcomes when considered together, individually by program, or for specific subgroups. For the combined-program analyses, the year-by-year analysis found fewer significant impacts than expected by chance (2 out of 60 estimated impacts) and the growth curve analysis found no significant impacts. For the individual program analyses, the year-by-year analysis found fewer significant impacts than expected by chance (16 out of 420 estimated impacts), with 9 having beneficial impacts and 7 having detrimental impacts. The growth curve analyses of the individual programs found the same number of impacts as expected by chance (6 out of 126 estimated impacts), with 2 having beneficial impacts and 4 having detrimental impacts. For the subgroup analyses, the year-by-year analyses found more significant impacts than expected by chance for gender (8 out of 54 estimated impacts) and initial risk levels (41 out of 270 estimated impacts). For gender, half of the significant impacts showed a beneficial impact of the intervention and half showed a detrimental impact. For initial risk levels, 26 showed a beneficial impact of the intervention on high-risk students and 15 showed a detrimental impact on high-risk students. The growth analysis found fewer significant impacts than expected by chance for all the subgroups except those based on initial risk levels (6 out of 90 estimated impacts). Four of the 6 significant impacts were beneficial and 2 were detrimental. In sum, the SACD combined-program evaluation provides no evidence that the seven universal, schoolwide programs improved students' social and character development.

Several explanations for this finding can be considered: (1) failure of the conceptualization and design of the intervention, (2) weak implementation of the intervention, (3) nonsubstantial differences in the level of SACD activities in the treatment and control schools, and (4) methodological limitations of the evaluation.

Failure of the conceptualization and design of the intervention refers to the possibility that the seven programs tested might not have altered students' social and character development in the expected ways because the theories underlying them or the combinations of activities chosen to bring about the desired changes in students' attitudes and behaviors were inadequate for the purpose. For example, one alternative view to that adopted in the SACD evaluation is that only a subset of elementary-aged children has deficits in social behavior and character, and these deficits require a more targeted, more intensive intervention than schoolwide programs can provide. Therefore, for a school-based program to be effective, a combination of schoolwide and targeted activities might be required for the intervention to make a significant difference in student outcomes (e.g., see Conduct Problems Prevention Research Group 1999). Such explanations, if confirmed by other studies, would lead toward focusing more effort on understanding how social and character development occurs among elementary-aged children, how this development can be affected, and what types of practices in classrooms and schools can be used to bring about desired effects.

Weak implementation of the intervention refers to the possibility that, although the intervention might have been well conceived and well designed, the treatment schools did not implement the practices effectively on average. Weakly implemented programs may not have positive impacts on students. The SACD evaluation examined seven different SACD programs, each with unique features, and the fidelity rating that was used measured fidelity of implementation relative to the targets established for each specific intervention. In this

way, the ratings were standardized relative to each site's program-specific benchmarks, allowing them to be compared across programs and years. The analysis of the fidelity data found little evidence of a relationship between high fidelity and more beneficial outcomes. The number of significant associations found between fidelity and beneficial outcomes was higher than expected by chance (5 found, with 3 expected by chance out of 54 estimated impacts), but 4 of the 5 significant results were due to detrimental associations between low fidelity and outcomes rather than beneficial associations between high fidelity and outcomes. The approach used to obtain comparable ratings required two compromises. First, it could not account for differences among the programs' implementation standards—for example, whether programs differed in how difficult they were to implement. Second, it provided little information about why implementation was of a certain quality. As a result, the SACD evaluation fidelity measure may not provide adequate information about whether low fidelity might have been the reason behind the lack of significant findings. The fidelity measures used by each research team, which used team-chosen criteria, may provide additional information on how well each program was implemented in each treatment school.

Nonsubstantial differences in the level of SACD activities refers to the possibility that the implementation differences between the treatment and control schools were not great enough to generate statistically significant differences in student outcomes. Like the treatment schools, the control schools joined the study with a willingness to implement a SACD program showing a willingness to promote social and character development. In addition, some of the sites were located in states where legislation required or promoted such activities. The control group represented "standard practice," which included the reported use of SACD activities in the classroom. For example, 86 percent to 90 percent of control teachers reported using activities to promote any one of the six SACD goals. While a statistically significant larger percentage of teachers in the treatment schools (95% to 96%) reported conducting such activities, the 5- to 10-percentage-point differences may not have been large enough to lead to improved student outcomes. At the same time, the significant differences between treatment and control teacher reports were larger than 5 to 10 percentage points for other responses regarding the use of SACD activities. For example, the differences between treatment and control teachers regarding the use of activities from specifically named programs were 29 to 34 percentage points across the 3 years. These results, plus the finding that treatment teachers reported greater use of some instructional materials and methods to promote SACD goals, provide evidence that the treatment teachers were making a more intensive effort to promote social and character development.

There are three methodological limitations of the evaluation that may have contributed to the finding of no impacts on student outcomes.

First, the evaluation relied on self-reported data by teachers and principals regarding the use of SACD activities. Observational studies were not done to validate these reports. If treatment teachers over-reported their use of SACD activities (possibly because they felt an expectation to report high use given that a SACD program was being implemented in their school), the impacts of the treatment could be misestimated. That is, if there really were no differences in the levels of SACD activities between the treatment and control groups, then a lack of effects might be expected. However, treatment teachers did differ in their reported use of any activities to promote one of the six SACD goals (95% to 96% reported such use) versus their reported use of any activities from named programs (68% to 72%) This, while not ruling out the possibility of systematic over-reporting, might suggest that some teachers were candid in their reporting on their use of the treatment programs. The research teams used observations or product measures to check implementation of their specific programs, and the results from this work may provide additional evidence about the potential importance of over-reporting of implementation by treatment teachers.

A second methodological limitation was that student-provided data (used for 12 of the 20 outcomes) were not available for 36 percent to 39 percent of students, depending on the year, because primary caregivers did not provide written consent or students did not assent to take part in the study (primary caregiver data, used for three outcomes, were not available for 49% to 54% of students). It is possible that students included in the study differed from those not included due to an absence of data. As the study did not collect descriptive

data on the nonobserved students, the existence of such differences could not be determined, and how the inclusion of these students in the study would have affected the findings is unknown. Given the few statistically significant and substantively important impacts found with the existing sample, there would need to be a large and consistent impact on the nonobserved students (had they been included) to change the findings. For example, because two-thirds of the population who were observed received an average impact of zero, the nonobserved one-third would need to have received an average impact of nearly one-third of a standard deviation to bring the overall mean to one-tenth of a standard deviation unit. In addition, because the subgroup analyses did not find systematically significant impacts, there is no evidence that should these missing students come from one of the subgroups (e.g., higher initial risk) their inclusion would change the findings.

A third methodological limitation was the sample size for the individual program evaluations. The combined analysis of all seven programs provided a sample size sufficient to detect student-level impacts at minimum detectable effect sizes (MDES) ranging from 0.03 to 0.23 standard deviations (with more than 75% of them below 0.10 standard deviations). The power to detect impacts at the level of the individual programs was more limited (the MDES ranged from 0.09 to 1.04 over the 3 years), and individual program-level effects might have been missed. To address this limitation at the program level, nonsignificant impacts of at least 0.25 standard deviations were identified as substantively important results. Only a small number of these were identified, with an almost equal division into beneficial and detrimental effects.

The seven programs tested in the SACD evaluation were similar in being coherent, universal, school-based programs to promote social and character development of third- to fifth-grade students. They were diverse both in their specific goals and in their approaches to promoting social and character development for all students. In addition, they were evaluated in diverse types of locales in which schools served communities with very different ethnic and socioeconomic compositions. On average, the seven programs did not improve students' social and emotional competence, behavior, academic achievement, and student and teacher perceptions of school climate. In addition, although the numbers of schools and students in each program were not always sufficient to support firm conclusions at the program level, the patterns of estimated impacts for each program were largely similar: students' outcomes were not affected.