Journal of Advanced Academics
jaa
Volume 18 ✣ Number 1 ✣ Fall 2006 ✣ pp. 82–114

Relationships Between Implementing Character Education, Student Behavior, and Student Achievement

Gary Skaggs
Nancy Bodenhorn
Virginia Polytechnic Institute and State University

One of the requirements of the Department of Education, through the No Child Left Behind Act, is to conduct rigorous, experimental design studies showing the impact and results of educational endeavors. Character education (CE), while widely implemented in schools nationwide, has not been widely researched using comparative studies. Studies have been conducted, indicating positive student development resulting from CE programs or programs that provide a desirable CE environment (Harrington, Giles, Hoyle, Feeney, & Yungbluth, 2001; Leming, 2000; Schultz, Barr, & Selman, 2001; Williams, Yanchar, Jensen, & Lewis, 2003). However, many programs available for purchase or implementation remain either: (a) evaluated only by internal evaluators, (b) not scrutinized by an academic review process, (c) evaluated through a grant process that frequently becomes the property of the sponsoring agency, or (d) unevaluated. As a result, school administrators, the con-

Summary

Over a 4-year period, researchers measured several outcomes in 5 school districts initiating or enhancing character education programs. Based on student, teacher, and administrator surveys, there was a noticeable improvement in character-related behavior. In certain districts, suspension and drop-out rates also decreased after the implementation of the character education programs; however, the relationship between these behavioral indicators and character education was inconclusive. Character education programming had little impact on student achievement, perhaps because of the lack of a direct relationship between character education goals and student achievement goals. In addition, the research examined the relationship between the implementation level of the program within the school and the measured outcomes. As expected, schools with more fully-implemented programs experienced a greater improvement in perceived character-driven behavior and lower suspension rates than schools with less well-implemented programs. The highest degree of implementation in schools resulted when personnel at those schools embraced the program as their own. This underlined the necessity for gaining community and staff support for the implementation of a character education program.

Skaggs, G., & Bodenhorn, N. (2006). Relationships between implementing character education, student behavior, and student achievement. *Journal of Advanced Academics, 18*, 82–114.

sumers of these products, rely on either anecdotal information and/or information provided by the product developers when deciding on a CE program for their students.

Character education was a major part of the public schools' mission until the 1950s, when it was phased out of public education. A major reason for its decline was a fear that teaching morality would be equated to the teaching of religion. Character education made a comeback in the 1980s and 1990s as a result of perceptions of the decline in the quality of public education (Lickona, 1991). Many nonacademic reform efforts have been undertaken to reverse this perceived decline, including community service, peer mediation, drug prevention, and CE.

Good character is generally described as involving the facility to consistently apply principles such as respect for others, truthfulness, fairness, and responsibility when facing behavioral and ethical choices (Lickona, 1991). Character education is described as "any deliberate approach by which school personnel, often in conjunction with parents and community members, . . . help children and youth become caring, principled, and responsible" (Williams, 2000, p. 32). A variety of CE programs is now available. A recent visit to the Education for Character Web site (http://www.cortland.edu/character/chared_orgs.html) listed 41 programs or organizations offering some form of CE. Some programs target specific aspects of character development, such as bullying, communication skills, or community building, while others are more comprehensive. The current trend in character education practice involves an emphasis on similarities rather than differences; a balance of moral reasoning skills and establishing behavioral habits; a balance of focus on the responsibilities of the individual and the community; an inclusion of cognitive, affective, and behavioral dimensions; and an expectation for educators to serve as role models for students (Williams, 2000). DeRoche and Williams (2001) reviewed effective CE programs and found the following common components: vision, standards, expectations, implementation criteria, leadership, resources, training, partnerships, and assessment.

This study grew out of an evaluation of CE programs initiated in 5 public school districts in a state in the eastern United States. The evaluation resulted from a U.S. Department of Education (DOE) grant to assist states in implementing new CE programs or supplementing existing CE programs. The grant project was called the Partnership in Character Education and was comprised of the state department of education and 5 school districts, referred to as partner schools and labeled in this study as Districts 1 through 5.

The purpose of this study was to explore the results of implementing CE programs in general and the impact of implementation level specifically. The design was not experimental, which limits causal inference that can be made regarding the effects of CE. This study does, however, offer a unique opportunity to examine CE in a large number of schools and in a wide variety of settings. This study covers a 4-year period from program planning through the first 3 years of implementation.

Comparing programs in different school districts would not be effective in evaluating a specific approach to CE because, although districts adopted formal programs, each district administered the programs in different ways. As a result, this study focused on exploring the relationship between the presence of CE and outcome measures, and was guided by the following research questions:

1. Is there a relationship between the presence of a character education program and perceptions of student and staff behavior, behavioral indicators, and school achievement?
2. Is there a relationship between the degree of implementation of character education and measures of perceptions of student and staff behavior, behavioral indicators, and school achievement?

Character Education Programs

Each of the 5 school districts used CE curricula from one of four formal programs. The programs and implementation pro-

Table 1

Description of Character Education Programs in Five School Districts

District	Description of District	Character Education Curriculum	Administrative Structure	Schools in study (*n*)
1	Rural/exurban, middle class	Educating for Character	Program determined by school improvement teams	18
2	Urban inner city, large minority, low SES	Character Education Inst./Community of Caring	Program run by central administration	12
3	Suburban, middle class	Educating for Character	School program determined by committee of parents, students, and school staff	13
4	Suburban, large minority, low to middle class	Community of Caring	School program determined by school staff, approved by central admin.	12
5	Suburban/exurban middle class	Character Counts!	School staff trained by Josephson Inst.; all schools implement program	49

cesses are described below. Table 1 summarizes the brief description of the districts, their structure and program curricula, and the number of schools that implemented CE.

District 1

District 1 contains 3 high schools, 4 middle schools, and 11 elementary schools. Located outside a major metropolitan area, it was a rural community but now is a location of intense suburban development. The district had been using school improvement action teams for several years before implementing CE. The intent in District 1 was to integrate CE into the school improvement structure. Curricular materials for CE were developed locally based on Lickona's (1991) *Educating for Character*. This program, which focuses on "developing performance char-

acter (doing our best work) and moral character (doing the right thing) within an ethical learning community" (Center for the 4th and 5th Rs, n.d., ¶1), utilizes strategies within the school and the broader community to impact awareness, attitudes, and action. Their comprehensive strategy includes role modeling, creating a caring and democratic classroom community, character-based discipline, cooperative learning, ethical reflection, conflict resolution skills, and integrating character throughout academic curricular lessons (for more information, please visit http://www.cortland.edu/c4n5rs).

District 2

District 2 is an urban system, containing 171 schools. A CE program was implemented prior to the partnership grant. Elementary and middle schools used materials developed by the Character Education Institute of San Antonio, TX. Although the institute is no longer operational, the program emphasized societal values such as honesty, patriotism, generosity, and responsibility. At the high school level, the district used the Community of Caring model. The Community of Caring program emphasizes the inclusion of students with disabilities and community service. It focuses on the five core values of caring, respect, responsibility, trust, and family. This program utilizes family networks and a variety of school forums in which students are encouraged to participate and develop leadership (for more information, please visit http://www.communityofcaring.org). Grant funds supplemented the existing programs in 4 high schools, 4 middle schools, and 4 elementary schools with teacher and parent in-services and additional curricular materials. This study included data from these 12 schools.

District 3

District 3 is a suburban district with 152 schools. During the early 1980s, District 3 developed its own CE program. At the time of this project, the district decided to change its program

to the Educating for Character program. Grant funds were used primarily for 4-day intensive summer institutes, during which the staff from a small number of schools was trained in how to implement the program and train personnel in their schools. Institute participants then applied what they learned in their schools. Data from the first group of 11 elementary and 2 middle schools were used in this study.

District 4

District 4 is a suburban district containing 163 schools. Using grant funds, the system initiated the Community of Caring program in 8 elementary, 2 middle, and 2 high schools. By the end of the second year of the grant, every school in District 4 had adopted or begun training in the use of some form of CE. In addition, every middle and high school adopted a peer mediation program. However, this study included data only on the original 12 schools that received grant funds.

District 5

District 5 is also a suburban district containing 9 high schools, 10 middle schools, and 30 elementary schools. This district implemented the Josephson Institute's Character Counts! program in all schools. The Character Counts! program centers on ethical decision-making in the context of the six pillars of character: trustworthiness, respect, responsibility, fairness, caring, and citizenship. These six pillars, or values, are incorporated into all aspects of the students' school life, including content areas, playground activities, and cafeteria behavior (for more information, please visit http://www.charactercounts.org).

Despite the variety of CE curricula and program administrative structures, the experiences of the students in each of the schools were remarkably similar. In most schools, there was a focus on a particular value or virtue each month. This value was incorporated into regular classroom instruction, and materials were sent home to parents. In addition, special events, such as a

values fair, were held and special displays were shown in hallways, on bulletin boards, and in gathering areas such as the cafeteria. As a result, this study is not about a specific approach but about the effect of implementing a CE program that is appropriate for a given community and the degree to which it is implemented.

Method

Design

This study is best considered a longitudinal panel study. Beginning with a baseline year, 1996–1997, before program implementation, it continued through 3 years of program implementation, 1997–1998, 1998–1999, and 1999–2000. The original plan for the evaluation was a quasi-experimental design for the 3 school districts (Districts 2, 3, and 4) that implemented CE in a limited number of schools. The intent was to obtain student behavioral and achievement information from similar nonprogram schools in these districts as a form of treatment versus control group comparison. However, by the end of the second year, it became clear that any such analyses would not be internally valid. All of the comparison schools had adopted some form of CE during the project or adopted materials from the participating schools. As a result, the comparison group component of the design was dropped.

By the end of the study, in 2000, only 7 districts in the state had not formally begun programs in CE. Achievement data and dropout rates from schools in the 7 nonprogram districts were obtained from the state department of education database and used as a comparison group. This comparison group was not ideal; the districts were not similar in many respects to the 5 partner districts. The 7 nonprogram districts were either midsized suburban or small, rural districts. However, after controlling for a number of demographic variables, we examined differences in trends between the two groups of districts.

Measures

Because CE focuses globally on character-related behavior rather than specific behavior, there is no universal set of outcome measures. The goals for programs are usually stated in general terms: character-driven behavior for Character Counts!, caring school community for Community of Caring, and moral school culture for Educating for Character. On the other hand, the Web sites for these programs describe program evaluations that consistently include other outcome variables such as student achievement, student behavior (i.e., risk behaviors, behavioral referrals, suspensions, dropouts, and the like), and perceptions of student behavior (i.e., school climate, student attitudes, self-esteem). During this study, school administrators consistently hoped that CE would improve both student behavior and relationships between and among students and school staff. District and state department staff valued improved student behavior, but also saw it as a means to an end, namely improved student achievement. As a result, data used included: (a) responses to a survey on perceived character-related behavior, (b) behavioral indicator data supplied by school personnel or the state department's database, and (c) school-level achievement information from the state department's database.

Behavioral Perceptions. The School as a Caring Community Profile (SCCP; Lickona, 1995) consists of 30 Likert items that ask community members (students, staff, and parents) about their perceptions of community members' behavior. The first 20 items measure perceptions of student behavior. Sample items include "Students treat classmates and schoolmates with respect" and "Students refrain from put-downs." Items 21–28 evaluate perceptions of teacher and school staff behavior, and include items such as "In their interactions with students, teachers display the character qualities" and "Teachers treat all students fairly and don't play favorites." The last two items on the SCCP are related to perceived parent behavior; they will not be reported separately because there are only two items. Total scale and subscale means were used for analysis, as opposed to sum scores. Scale means

can range from 1.0 to 5.0, with higher means indicating higher levels of positive behavior that are character-related.

For the target population in this study, to verify the internal consistency reliability of the SCCP, coefficient alphas were calculated for each respondent group (administrators, teachers, students) using the baseline year data. They were calculated separately for the student and school staff subscales. For the student behavior items, alphas ranged from .93 to .97. For the school staff items, alphas ranged from .89 to .93.

Behavioral Indicator Data. Initial behavioral indicators included suspensions, expulsions, attendance rates, and dropout rates. Each year, administrators from each partner school provided information on the number of suspensions and expulsions. Suspension rates for each school were calculated by dividing the number of suspensions by the school's enrollment and multiplying by 100. Because expulsions were very rare, they were not used in the analysis. Dropout rates were supplied from the state education department's database. Attendance rates were also available from this database. However, attendance rates were very high across the state and relatively stable from year to year, and thus attendance was not used in this study.

Student Achievement. Student achievement information came from the state assessment, administered to grades 3, 5, and 8 in a variety of subject areas. This study examined fifth- and eighth-grade reading and mathematics scores. The high school assessment at that time was a series of minimum competency tests for which the average pass rates ranged from 95% to 99%. There was too little variability in the high school scores to use this assessment for analysis. The state department provided the assessment data in the form of percent of students achieving a "satisfactory" level.

Procedure

In the spring of each year, from the baseline year to the third year of program implementation (1997–2000), administrators, teachers, students, and parents in the 5 partner school districts

completed the SCCP. In each partner CE school, all teachers and administrators were administered the survey; the response rates for these groups averaged more than 90%. The survey was also administered to students in randomly chosen classrooms. In elementary schools, these classrooms were restricted to the fifth grade. These students were then given the survey to take home for their parents to complete. In District 4, between 5% and 10% of the total number of students in the partner schools completed the survey. In Districts 1, 2, and 3, between 10% and 20% of the students completed the survey. In District 5, 15–30% of the students completed the survey. In District 5, a random sample of one fourth of the middle and high schools was administered the surveys. Across all 5 districts, less than 10% of the parent surveys were returned. Therefore, separate results for parents are not reported; however, they were included in the total sample. The total number of respondents was 32,498 in the baseline year, 17,770 in the first program year, 16,984 in the second year, and 17,025 in the third year.

Results

Behavioral Perceptions

Table 2 presents means and standard deviations of the total SCCP scores for the 5 districts. In 1997, the baseline year prior to implementation, the means ranged from 2.88 to 3.67. The lowest SCCP means were for Districts 2 and 4. In 1998, after one year of implementation, mean SCCP scores rose in all 5 school districts. These means ranged from 3.20 to 3.80. In 1999, the second year of implementation, SCCP scores dropped slightly from the previous year, with the exception of District 3 scores, which were still higher than the baseline year. In the third year of implementation, there was an overall slight rise close to the first implementation year level.

Cohen's *d* effect sizes are reported in the last column of Table 2. In the case of SCCP means, effect sizes indicate how many

Table 2

SCCP Descriptive Statistics

	1997	1998	1999	2000	Cohen's *d*
District 1					
Mean	3.39	3.64	3.58	3.65	
SD	0.76	0.76	0.75	0.73	0.36
District 2					
Mean	2.92	3.20	3.18	3.20	
SD	0.74	0.67	0.70	0.67	0.40
District 3					
Mean	3.67	3.80	3.60	3.75	
SD	0.65	0.61	0.67	0.66	0.12
District 4					
Mean	2.88	3.41	3.30	3.21	
SD	0.70	0.62	0.63	0.70	0.47
District 5					
Mean	3.20	3.60	3.51	3.56	
SD	0.77	0.72	0.76	0.73	0.48
All districts					
Mean	3.19	3.58	3.49	3.53	
SD	0.77	.072	0.74	0.73	0.45

Note. **Effect size is in *SD* units between 1997 and 2000.**

standard deviations apart the means were between the baseline and third year of implementation. In 4 districts, the effect sizes ranged from .36 to .48 of a standard deviation, a small to medium effect (Cohen, 1988). District 3 had a much smaller effect, only .12 standard deviation units. Figure 1 shows the total and student (items 1–20) and staff (items 21–28) item subset means across the 4 years for all 5 school districts combined. The results were very similar within each district. Each year, the means of the staff behavior items were higher than the means for student behavior items. The difference ranged from .70 to .80 standard deviation units, a relatively large effect size, meaning that respondents perceived staff behavior as being substantially more character-based than student behavior.

Figure 2 shows mean SCCP scores according to the type of respondent, combining data from all 5 districts. As in the previous figure, the relative ordering of means by respondent group

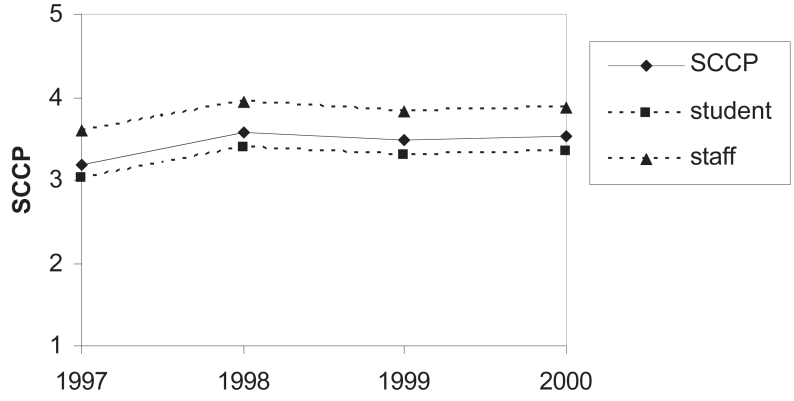

Figure 1. Mean SCCP for total, student, and staff item subsets

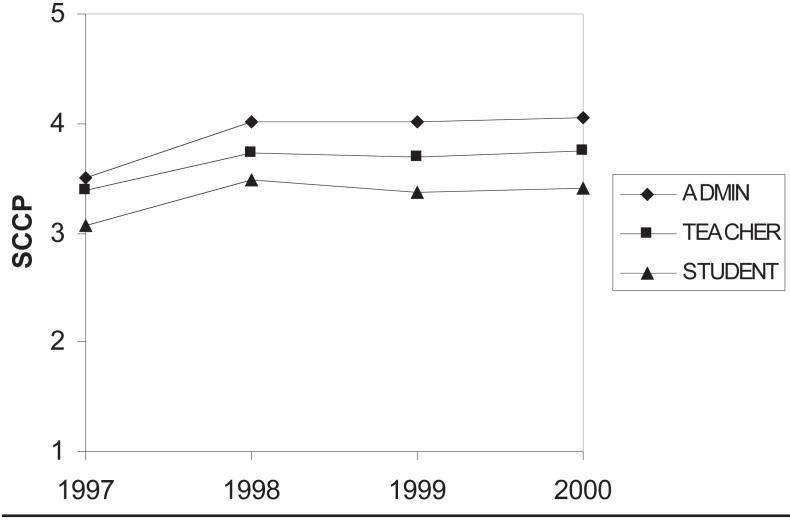

Figure 2. Mean SCCP by respondent group

was the same in the 5 districts. In all 4 years, administrators perceived the highest level of character-based behavior, followed by teachers. Students had the lowest perceptions of positive behavior. The difference between student and administrator perceptions was about half a standard deviation unit. In other words, students viewed behavior in school the least positively of all respondent groups, whereas administrators had the most positive perception. Teachers' perceptions were between those of

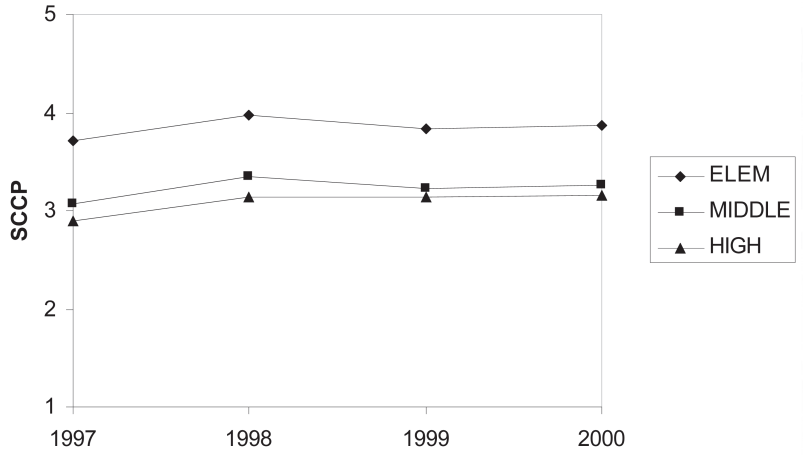

Figure 3. Mean SCCP by school level

administrators and students. All respondent groups perceived an improvement in perceived behavior after the baseline year. In Figure 3, SCCP means are shown according to school level. As in the previous figures, the results are for all districts combined. The relative ordering by school level was the same in each district. In all years, perceptions of elementary school behavior were much more positive than those of secondary school behavior, a difference that approached a full standard deviation. This is not surprising, as middle and high schools are considered more difficult places to manage student behavior than elementary schools (McWhirter, McWhirter, McWhirter, & McWhirter, 2004).

The SCCP means shown above represent an average across all survey items. It is also important to examine the behaviors represented in individual items. Table 3 shows the five most and least positive survey items for all 4 years of the study. The same statements appeared as most and least positive each year the survey was administered, by respondents from all school districts, all respondent types, and all school levels. The most positive statements all dealt with school, teacher, or staff behavior. The least positive statements were about student behavior. Table 4 lists those items that showed the most improvement from 1997 to 2000, in terms of both the raw score units and effect size (standard deviation units). These improvements ranged from about

Table 3

SCCP Means for the Five Most and Least Positive Items

Most Positive Items		1997	1998	1999	2000
Teachers respect, care about, and help each other.	Mean	3.85	4.17	4.05	4.08
	SD	1.13	1.01	1.05	1.04
The school treats parents in a way that makes them feel respected, welcomed, and cared about.	Mean	3.73	4.13	3.98	4.00
	SD	1.23	1.07	1.13	1.10
Teachers go out of their way to help students who need extra help.	Mean	3.65	4.06	3.91	3.97
	SD	1.25	1.08	1.12	1.10
In their interactions with students, other professional school staff displays the character qualities the school is trying to teach.	Mean	3.80	4.09	3.96	3.97
	SD	1.17	1.07	1.10	1.09
In their interactions with each other, staff displays the character qualities the school is trying to teach.	Mean	3.76	4.04	3.92	3.95
	SD	1.16	1.06	1.09	1.08
Most Negative Items					
Students care about and help each other, even if they are not friends.	Mean	2.67	3.05	2.93	2.98
	SD	1.22	1.16	1.15	1.17
Students solve conflicts without fighting, insults, or threats.	Mean	2.51	2.97	2.88	2.96
	SD	1.24	1.24	1.22	1.22
When students do something hurtful, they apologize and try to make up for it.	Mean	2.48	2.95	2.89	2.94
	SD	1.21	1.22	1.19	1.19
Students refrain from put-downs.	Mean	2.46	2.88	2.79	2.86
	SD	1.21	1.20	1.17	1.16
When students see another student being mean, they try to stop it.	Mean	2.44	2.82	2.80	2.85
	SD	1.17	1.20	1.17	1.17

0.3 to 0.4 standard deviations. Interestingly, 8 of the 11 items referred to student behavior, including 4 of the 5 least positive statements that appeared in Table 3. This suggests that, although student behavior was rated relatively poorly, a significant portion of the perceived improvement came from improved student behavior.

Behavioral Indicators

Table 5 shows the means and standard deviations of suspension and dropout rates for the 5 partner districts and all CE

Table 4

Means for SCCP Items Showing the Most Improvement

Item	Change from 1997 to 2000	Cohen's *d*
When students do something hurtful, they apologize and try to make up for it.	0.46	0.38
Older students are kind to younger students.	0.46	0.38
Students solve conflicts without fighting, insults, or threats.	0.45	0.38
Teachers treat all students fairly and don't play favorites.	0.42	0.32
When students see another student being mean, they try to stop it.	0.41	0.35
Students refrain from put-downs.	0.40	0.35
Students help new students make friends and feel accepted.	0.39	0.35
Students are patient and forgiving with each other.	0.37	0.34
Teachers listen to students' problems, and students feel they can talk to their teachers about things that are bothering them.	0.37	0.28
Students refrain from picking on others because they are different.	0.35	0.29

schools combined. Dropout rates for the schools in the non-CE program districts are also shown. Suspension rates were generally low in elementary schools but much higher in middle and high schools. At the secondary level, there was also considerable variability in suspension rates between districts. The large standard deviations indicate substantial variability between schools, as well.

Figures 4, 5, and 6 show the suspension rates of the 5 partner districts and all districts combined for elementary, middle, and high schools, respectively. Across all districts and at all three levels, the suspension rate rose slightly in the first year of implementation and then decreased in each of the following years. Among the districts, District 4 had the highest suspension rates across the three levels but also the largest decline in suspension rates. Most districts had stable suspension rates across all 4 years. The exceptions were District 3 elementary schools, whose rates rose between 1997 and 2000, and District 2 middle schools,

Table 5

Mean Percentages of School Enrollment on Behavioral and Achievement Variables for Final Year of Study

	Elementary	Middle	High
Suspension Rate			
District 1	1.57 (1.6)	15.13 (6.7)	18.36 (4.6)
District 2	0.26 (0.2)	2.01 (2.4)	1.97 (1.9)
District 3	3.17 (2.7)	18.96 (12.6)	
District 4	2.19 (1.0)	19.28 (8.5)	36.37 (7.9)
District 5	1.23 (1.7)	22.25 (11.8)	30.91 (9.9)
All CE Schools	1.65 (1.9)	18.05 (11.9)	23.26 (14.3)
Dropout Rate			
District 1			3.13 (0.7)
District 2			5.14 (4.9)
District 3			
District 4			3.08 (0.2)
District 5			2.05 (2.7)
All CE Schools			3.03 (3.1)
Non-CE Schools			3.29 (4.4)
Reading (Pct. Satisfactory)			
District 1	55.56 (10.0)	34.56 (10.5)	
District 2	23.16 (5.7)	10.58 (9.9)	
District 3	47.26 (14.4)	33.58 (13.8)	
District 4	27.13 (11.0)	18.84 (4.6)	
District 5	53.03 (10.1)	28.66 (7.4)	
All CE Schools	46.82 (15.2)	23.76 (14.1)	
Non-CE Schools	43.31 (19.2)	25.47 (13.9)	
Math (Pct. Satisfactory)			
District 1	56.22 (12.9)	63.47 (12.6)	
District 2	24.64 (12.0)	18.60 (16.9)	
District 3	42.08 (17.4)	58.62 (18.6)	
District 4	24.39 (17.0)	23.70 (0.0)	
District 5	53.58 (11.5)	63.92 (7.7)	
All CE Schools	46.73 (18.2)	47.07 (26.0)	
Non-CE Schools	46.20 (23.4)	48.71 (23.4)	

Note. **Standard deviations in parentheses.**

whose rates fell. However, because these results were based on a relatively small number of schools, these findings are tenta-

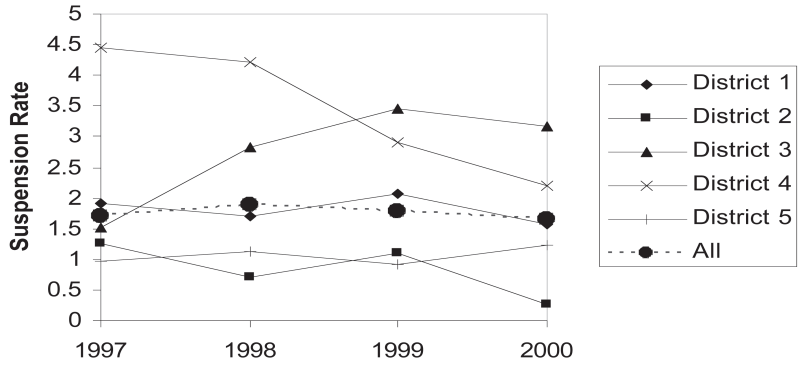

Figure 4. Suspension rates for elementary schools

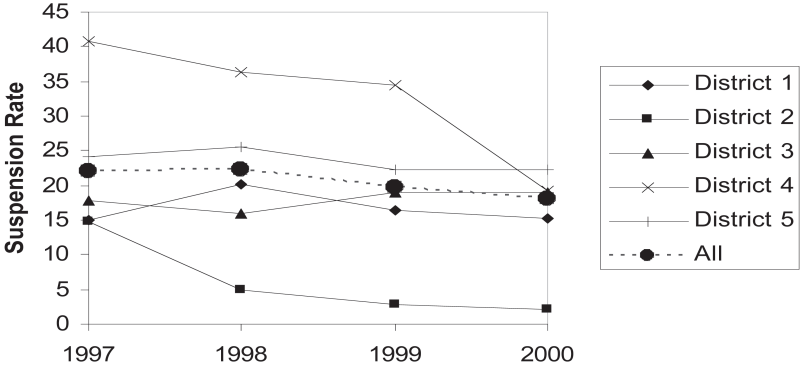

Figure 5. Suspension rates for middle schools

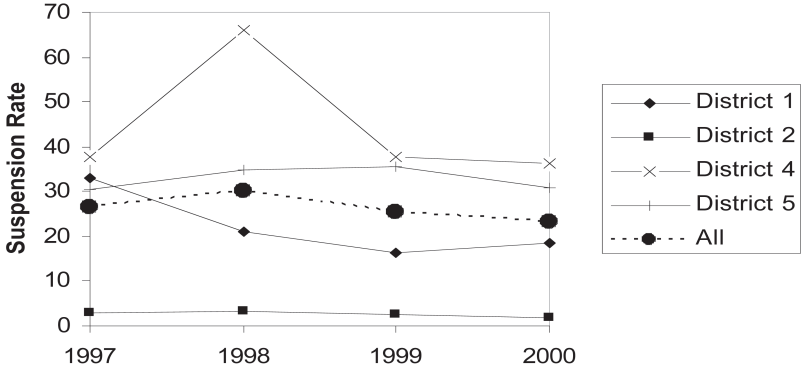

Figure 6. Suspension rates for high schools

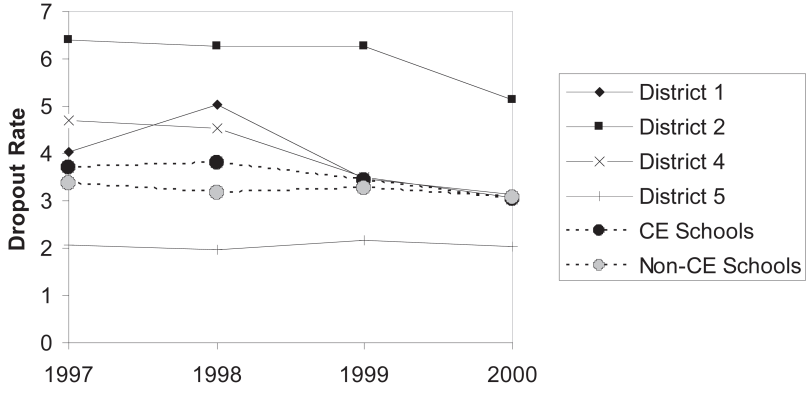

Figure 7. Mean dropout rates for high schools

tive at best. As a result, there is no clear relationship between suspension rates and the presence of CE. The one exception was District 4, which involved highly motivated schools.

Figure 7 shows the high school dropout rates for schools in the 4 partner districts (District 3 did not involve high schools), as well as the non-CE schools. There was an overall decline in the dropout rate over the 4-year period in all districts except District 5. In these 3 districts, most of decline occurred after 1998. We note though that the small number of schools involved (fewer than 5 per district) makes this finding very tentative. By contrast, in the non-CE schools, the dropout rate stayed virtually the same over the 4-year period. To control for some of the pre-existing differences between the CE schools and non-CE schools, a hierarchical multiple regression was carried out. The 2000 dropout rate was regressed on the 1997 dropout rate and percent of students receiving free or reduced lunches, Title I assistance, English as a second language instruction, and the CE versus non-CE school distinction. Schools in partner districts that were not supported by the grant were excluded from the analysis. Because schools were the unit of analysis, the regressions were weighted by the total school enrollment. The results showed that controlling for the 1997 dropout rate and the socioeconomic status variables, the partner versus nonprogram district grouping, or between schools with or without a CE program, accounted for only 0.6% of the variance. The adjusted mean dif-

ference in dropout rates was 0.4% lower for the partner districts. In other words, the relationship between CE and the dropout rate was virtually nonexistent.

School Achievement

The lower half of Table 5 provides means and standard deviations of the percentage of students who achieved a satisfactory level in the 5 partner districts and non-CE schools. Districts 2 and 4 are relatively low achieving; in fact, they are generally two of the lowest scoring districts in the state. Districts 1, 3, and 5 had considerably higher levels of achievement. However, there was also considerable variation in the achievement between schools within each district.

To explore the relationship between the presence of CE and student achievement, we conducted a series of hierarchical multiple regression analyses in which the school achievement measure in 2000 was the dependent variable. The independent variable was whether or not CE was implemented, that is, CE schools versus the non-CE schools. As in the dropout rate analysis above, schools in partner districts that were not supported by the grant were excluded from the analysis. Covariates included the 1997 achievement measure, the percentage of students who were Title I, the percentage of non-native English students, and the percentage of students who received free or reduced lunch. Finally, variables to measure the interaction of the independent variable (CE schools versus non-CE schools) were created by multiplying the 1997 achievement (percent satisfactory) by a dummy variable (CE or non-CE school) and added to the regression analyses. The interaction variables estimated the degree to which the achievement differences between CE and non-CE schools varied according to their initial school achievement levels. Because the enrollment varied widely between schools and between school districts (urban schools tended to be larger), all analyses were weighted by the 2000 school enrollment as reported in the state database. The intent of the weighting was

to reduce school size as a confounding factor in the analyses and facilitate comparisons between the regression analyses.

The results of these regressions are shown in Table 6. For each grade and subject area, both the unstandardized and standardized regression coefficients are shown, as well as the change in multiple *R*-squared. The unstandardized coefficients represent the mean differences between CE and non-CE schools in terms of the percentage of students who achieved satisfactory status on the state exam, adjusted for the covariates. The standardized coefficients indicate the number of standard deviation units apart the two groups of schools were on achievement after controlling for all of the other variables in the model. Positive coefficients indicate that schools with CE programs had higher achievement than the non-CE schools. Negative coefficients indicate higher achievement for non-CE schools. Because of the small number of middle schools, only the overall coefficients are shown for eighth grade.

All standardized coefficients had absolute values less than .30, indicating small effect sizes for both subjects and grades. Except for District 2, all of the overall coefficients were negative by less than two tenths of a standard deviation, indicating little practical difference in achievement between partner schools and schools in districts without formal CE programs. In District 2, the unstandardized regression coefficient was 13.91, indicating that 13.91% more students achieved mastery in the CE schools than in schools in districts without formal CE programs. This could seem to indicate substantial achievement differences, but the variability between schools was large enough to cloud this overall interpretation. The changes in multiple *R*-squared were 0.05 or less for all analyses. This indicates that the presence of a CE program or its interaction with initial achievement accounted for no more than 5% of the variance in 2000 achievement. The standardized interaction coefficients for eighth grade were near zero, suggesting that the difference between partner and non-CE schools did not vary according to 1997 achievement levels. On the other hand, for fifth-grade reading and math, most of the interaction coefficients were larger and positive. This meant that for schools with lower 1997 achievement levels, non-CE schools

Table 6

Hierarchical Regression Analyses for School Achievement and Partner Schools Versus Nonpartner School Districts

Grade 5		Coefficient (Unstandard.)	*SE*	Coefficient (Standard.)	R^2 Change
Reading					
District 1	Main Effect	-1.40	0.13	-0.03	0.00
	Interaction	0.03	0.01	0.03	0.00
District 2	Main Effect	7.61	0.24	0.10	0.01
	Interaction	1.19	0.02	0.44	0.04
District 3	Main Effect	-0.36	0.16	-0.01	0.00
	Interaction	0.78	0.02	0.48	0.01
District 4	Main Effect	-10.77	0.17	-0.17	0.02
	Interaction	0.27	0.02	0.98	0.00
District 5	Main Effect	-4.73	0.09	-0.14	0.02
	Interaction	-0.14	0.01	-0.20	0.00
Partner v.	Main Effect	-3.66	0.07	-0.12	0.01
Nonprogram	Interaction	0.11	0.01	0.16	0.00
Mathematics					
District 1	Main Effect	0.63	0.18	0.01	0.00
	Interaction	-0.04	0.02	-0.03	0.00
District 2	Main Effect	13.91	0.34	0.14	0.01
	Interaction	1.02	0.01	0.48	0.05
District 3	Main Effect	-8.68	0.22	-0.12	0.01
	Interaction	0.49	0.02	0.37	0.01
District 4	Main Effect	-11.53	0.24	-0.15	0.02
	Interaction	0.82	0.04	0.30	0.00
District 5	Main Effect	-6.56	0.11	-0.16	0.03
	Interaction	0.19	0.01	0.30	0.00
Partner v.	Main Effect	-4.92	0.09	-0.13	0.02
Nonprogram	Interaction	0.20	0.01	0.31	0.01
Grade 8					
Reading					
Partner v.	Main Effect	-4.68	0.07	-0.19	0.03
Nonprogram	Interaction	0.00	0.01	0.00	0.00
Mathematics					
Partner v.	Main Effect	-4.09	0.06	-0.11	0.01
Nonprogram	Interaction	-0.17	0.00	-0.03	0.00

had higher achievement 2000 scores than the CE schools. For schools with high 1997 achievement levels, CE schools had higher achievement scores. This interaction is shown in Figure

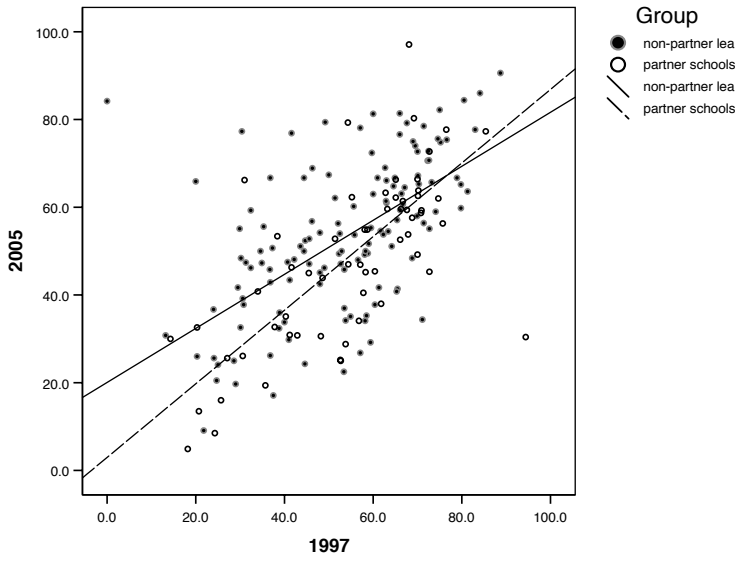

Figure 8. Fifth-grade mathematics achievement of CE and non-CE schools

8 for fifth-grade mathematics as separate regression lines for the CE and non-CE schools. The interaction for fifth-grade reading is similar but smaller, meaning that the regression lines have more similar slopes. These effects are small enough to make this finding extremely tentative.

The overall conclusion from these analyses is that there was little relationship between school achievement and the presence of a CE program. The differences between the two groups were small when controlling for 1997 achievement and several covariates related to socio-economic status. Furthermore, these small differences were not related to school initial achievement at grade 8. At grade 5, there was a tendency for the CE schools with higher initial achievement levels to have higher 2000 achievement levels than non-CE schools, and vice-versa.

Degree of Implementation

It is unrealistic to expect that each school within a district will implement an educational program exactly as prescribed

or to the same degree. Character education was introduced as a requirement for all schools in Districts 1 and 5. In District 4, pilot schools were chosen, mostly on their expressed interest in adopting a CE program. In Districts 2 and 3, schools were selected based on school interest and central office perception of need. In all districts, the formal program was chosen by the district and not by the schools themselves. For this study, there was no uniform way to measure the degree of program implementation. Instead, a group of "high implementation schools" was identified. These were schools that won or were nominated for awards for their programs or identified by school district staff as having exemplary programs. In District 4, all schools were enthusiastic about CE, and they had equivalent records, thus they were all designated as high implementation schools. Altogether, there were 16 elementary schools, 8 middle schools, 6 high schools, and 1 vocational center in this group.

Outcome data from the high implementation schools were aggregated and compared to the remaining CE schools. A series of hierarchical regression analyses examined the impact of implementation. These analyses were similar to those described previously in that the three socioeconomic variables were used as covariates along with 1997 achievement. The additional independent variable was the implementation level of the school. The dependent variables included the SCCP school mean, school suspension rate, and fifth- and eighth-grade reading and math achievement levels (percent satisfactory). In each analysis, interaction variables were created as the product of 1997 achievement and the independent variable (high implementation CE schools versus other CE schools). As above, all analyses used total school enrollment as a weighting factor.

The results of these analyses are shown in Table 7. A positive coefficient here indicated a higher mean on the outcome variable for the high implementation schools, adjusted for the covariates, whereas a negative value favored the remaining schools. The largest positive value was for the SCCP at the high school level. The largest negative values were for suspension rates at the elementary and middle levels, indicating lower suspension rates

Table 7

Standardized Regression Coefficients for High Implementation Schools

Assessment		Coefficient (Unstandard.)	*SE*	Coefficient (Standard.)	R^2 Change
SCCP					
High implementation	Elem.	-0.09	0.00	-0.14	0.01
Interaction		0.10	0.01	0.50	0.00
High implementation	Middle	0.02	0.00	0.06	0.00
Interaction		0.60	0.02	4.37	0.03
High implementation	High	0.08	0.00	0.22	0.02
Interaction		0.10	0.01	0.73	0.00
Suspensions					
High implementation	Elem.	-1.24	0.03	-0.26	0.04
Interaction		-0.29	0.01	-0.25	0.02
High implementation	Middle	-6.07	0.14	-0.22	0.04
Interaction		-0.65	0.02	-0.64	0.03
High implementation	High	1.48	0.08	0.04	0.00
Interaction		0.13	0.01	0.11	0.00
Reading					
High implementation	Elem.	-1.44	0.12	-0.04	0.00
Interaction		0.25	0.01	0.20	0.01
High implementation	Middle	0.89	0.10	0.03	0.00
Interaction		0.13	0.01	0.14	0.00
Mathematics					
High implementation	Elem.	-0.52	0.14	-0.01	0.00
Interaction		0.10	0.01	0.10	0.00
High implementation	Middle	-1.66	0.08	-0.03	0.00
Interaction		-0.17	0.00	-0.18	0.00

for the high implementation schools. All other coefficients were near zero. All *R*-squared changes were less than 0.05.

The coefficients for the interaction variables are also shown in Table 7. For the SCCP, these were positive and relatively large at all three levels. In practical terms, this meant that the high implementation schools had higher 2000 SCCP scores than low implementation schools when their baseline SCCP was

also high. In other words, the greater benefit toward character-related behavior of a high level of implementation was evident in schools whose student and staff behavior were initially positive. This was also true of suspensions at the elementary and middle school levels. Suspension rates at the high school level were not related to the degree of program implementation. Degree of implementation also had little relationship to school level achievement results.

Discussion

In discussing the results of this study, we note that the degree to which changes in the outcome variables can be attributed to the presence of a CE program is limited by the fact that a randomized experiment was not feasible. These findings instead report relationships between CE and outcome variables that warrant further research. In this study, rather than comparing programs, we examined whether the implementation of a CE program and the degree of implementation of that program were related to a series of widely used outcome variables.

According the Web sites of the CE programs used by the school districts, a general goal of CE programs is for students to demonstrate positive behavior that exemplifies character traits, such as respect, trust, and responsibility. On the other hand, when presenting research findings or "success stories," findings are frequently expressed in terms of higher achievement, fewer suspensions or behavioral referrals, and fewer risk behaviors (e.g., teen pregnancy, drug use). The dramatic results in some schools are expressed in terms of behavioral and achievement outcome measures, but they are not the stated purpose of most CE programs. The impact of character-related behavior on the more prominent outcome measures is implied but not directly stated. By contrast, other recently-introduced programs focus on specific behaviors, such as bullying and conflict resolution. For these programs, expected outcomes are clearly articulated, but for

CE, the ambiguity of determining what outcome is appropriate makes it difficult to isolate the impact that CE has in schools.

In this study, three types of outcome variables were examined: behavioral perceptions, behavioral indicators, and school achievement. In a sense, these variables represent different relationships to the stated goals of the CE programs. The items on the SCCP reflect perceptions of the types of student, teacher, and staff behavior that CE promotes and are fairly direct indicators of the desired outcome of CE. Behavioral indicators, such as suspensions and dropouts, reflect problem situations that officials hope will improve as a result of increased instruction about character-related behavior. These indicators are a less direct measure of the impact of CE because they are also influenced by other issues in the school, such as administrative policy. Finally, the bottom line in educational reform is to improve student achievement. Educators and parents hope that in schools with more positive, and less disruptive, behavior, students will achieve at a higher level. However, there are many other factors that influence a school's achievement level, including changes in staff, staff training, resources, changes in curriculum standards, and changes in the community.

Not surprisingly, the results of this study show a demonstrable relationship between CE and behavioral perceptions, mixed results for behavioral indicators, and no relationship with student achievement. As the outcome measures become further removed from the actual CE program, their relationship to CE weakens. As indicated by DeRoche and Williams (2001), expectations and assessments should be aligned with the implementation process.

For perceived character-related behavior, as measured by the SCCP, there was an improvement in all districts at all levels from all types of respondents in the first year of program implementation. This improvement was largely maintained through the subsequent 3 years. The observed effect size and its observation in all districts by all respondent groups and at all levels suggest that an authentic change took place. Of course, it is possible that if the survey had been administered in non-CE schools, they would

have shown a similar improvement, as well. However, there was no obvious reason, such as a statewide policy change or seminal event, for an expectation that this would occur.

In the first year, a great deal of enthusiasm was generated among staff about the programs. It was therefore not surprising to see an improvement in perceived behaviors. This is likely similar to "vision" as referred to by DeRoche and Williams (2001) in their assessment of effective programs. There was a slight dip in the second year followed by an improvement in the third year. Some school officials attributed the second year dip to the shootings at Columbine, which unfortunately occurred just a few weeks before the survey was administered. By the third year, the programs were not so new, and other instructional and behavioral programs may have been initiated, as well. Nevertheless, perceptions of behaviors remained high, therefore the programs seemed to have produced some lasting effect. This is, perhaps, the most encouraging finding in this study for CE programs.

Administrators perceived character-related behaviors the most positively of all respondent groups, and students the least positively. All respondent groups had higher perceptions of school staff behavior than perceptions of student behavior. There is likely a wide range of interactions between students that school staff cannot observe. Elementary schools had the most positive level of perceived behaviors and high schools had the lowest. Character education coordinators in all 5 partner districts admitted that CE was more difficult to administer and promote in high schools. It is possible that the main features of character education programs—value of the month, character-related lessons, special fairs and recognitions—are not age-appropriate for high school students. Programs that focus on specific behaviors, such as conflict resolution, bullying, and peer mediation, may be more effective at that level.

Suspension rates for many individual schools fluctuated dramatically across the 4 years while the suspension rates for other schools were relatively stable. There could be many reasons for this, including change in administrators. When there is an administrative change at a school, there may be an increase in

suspensions as a new administrator attempts to make a disciplinary impact. That said, District 4's suspension rates declined across the 4-year period at all three levels. The reasons for this are not known. However, perhaps the high degree of motivation within the participating schools and the generally acknowledged high degree of implementation impacted the suspension rates in District 4.

High school dropout rates generally declined, but the differences between CE and non-CE schools were small. Similar to suspension rates, it is difficult to know the degree to which the presence of CE was related to dropout rates because other school efforts and programs may have impacted the results. A better indicator of student behavior would have been a more direct indication of the number of behavioral incidents, such as office referrals. Most schools, however, either did not collect this information regularly or did not collect it in the same way, which hindered our ability to make any direct comparisons among the schools.

There is little evidence to suggest a relationship between CE and school-level achievement. There was no noticeable improvement in achievement after CE was implemented. In fact, achievement in the partner schools lagged behind the non-program districts even after controlling for pre-implementation achievement and socioeconomic status. Because so many factors other than CE influence achievement, it is perhaps unrealistic to expect CE to show a direct achievement effect.

The issue of the extent that each partner school implemented CE is also related to how it was implemented. In Districts 1 and 5, the district's central administration directed that every school initiate a program. In those districts, some schools embraced CE; others did not. In District 4, individual schools volunteered, and all of the schools implemented CE enthusiastically. We also note that District 4 experienced the second highest increase in perceived behavior and a marked decline in suspensions and dropouts. District 3 experienced the smallest improvement in perceived behavior, and suspension rates actually rose in elementary schools. Possibly, the staff training implemented with the

grant funds did not result in much visible change over the already existing CE program.

Schools that were recognized as having highly implemented programs had more positive perceived behavior and fewer suspensions at the elementary and middle school levels than the remaining partner schools. We regard this as a promising but tentative finding, confirming the concepts of leadership, resources, and partnerships as common characteristics of effective programs (DeRoche & Williams, 2001). This also tentatively supports the finding that students in schools with more fully implemented school counseling programs reported higher grades, feeling better prepared for their future, and a more positive school climate (Lapan, Gysbers & Sun, 1997). At public meetings during the late 1990s, state and district staff frequently lauded the efforts of many of these schools that had shown dramatic improvement outcomes such as suspensions, dropouts, crime, fighting, and achievement. However, it is not clear that CE was the driving force behind the dramatic results. Other changes, such as change in administrative staff, changes in administrative policies, the introduction of other programs (e.g., Drug Abuse Resistance Education [DARE], peer mediation), and overall school climate, could have played a role, as well.

Each of the 5 districts chose the program that best fit its community and modified it to fit its needs. Although the programs differed in formal program distinction and how they were administered, the daily life of students in the participating schools was remarkably similar: lessons infused with CE lessons, student recognition for character-based behavior, and special community and school events centered on CE. As a result, this study attempted to determine if there is a global relationship of implementing a CE program with a variety of outcomes, and if so, examine whether the degree of implementation of the program also plays a role.

This study provided evidence that introducing CE was related to perceptions of improved student and school staff character-based behavior. Its relationship to behavioral indicators was mixed, and there was no apparent relationship to achievement.

Schools with highly implemented programs showed a higher level of perceived behavior and lower suspension and dropout rates. However, one clear need is for CE programs to be clearer about their expectations for measuring their success.

Because of the nonexperimental design, the above findings are tentative and need to be replicated by further research. Ideally, a situation in which schools could be randomly assigned to treatments or at least in which comparison schools were available would permit a greater degree of confidence in ascribing effects to character education. This was not possible in this study. Additionally, a consistent system for collecting behavioral data would greatly enhance comparisons across schools and districts. In such a system, behavioral indicators, such as office referrals for behavioral incidents and their resolution, would be collected in the same way in all schools. Such a system could permit an intensive examination of the stated outcomes of character education over an extended period of time. It will always be difficult in an experimental or quasi-experimental design to isolate the impact of CE programs on outcomes such as student achievement and suspension, attendance, or dropout rates because many other factors besides CE affect those outcomes. However, systematically collected data over a period of time can provide some indication of program impact.

Implications for Character Education Program Implementation

Administrators and school personnel who are considering implementing a character education program may wonder how to choose the best CE program for their community. The specifics of the program chosen for implementation in a school or school system may not be as important as the process of implementation. Each community may have differing needs and goals for character education, and fitting the program to the needs of the community and the skills and resources of school personnel will result in a more thoroughly implemented program, which appears to influence the degree of change in students'

behavior. In many ways, District 4 is a model program. Highly motivated schools introduced a program that fit within the culture of their communities. The failure of the comparison schools component of the evaluation was due to highly motivated teachers who viewed character education as a solution to a problem rather than an additional administrative task. In districts where the central office mandated character education, implementation was much more uneven. The high degree of implementation of some schools in these districts resulted from personnel at those schools embracing the program on their own.

The goals and purpose of implementing a CE program should be clearly articulated, both for the selection and the evaluation rationale. Further, these goals should be appropriate for the program. Implementing character education to raise student achievement is not realistic; however, implementing it to improve student behavior and the overall climate of the school is quite appropriate. At the same time, we recommend initiating a system for evaluating outcomes of CE programs. One high school in this study had developed a detailed database for entering office referrals for behavioral problems and actions taken over the 4-year period. Had this system been in place in all schools, a more direct indication of the effect of CE on student behavior would have been available. Unfortunately, many schools and districts define and use suspensions in different ways. A systematic, consistent way of collecting behavioral data across schools would provide administrators with valuable information on the progress of their efforts.

References

Center for the 4th and 5th Rs. (n.d.). *Respect and responsibility.* Retrieved December 11, 2006, from http://www.cortland.edu/character

Cohen, J. (1988). *Statistical power analysis for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

DeRoche, E., & Williams, M. (2001). *Educating hearts and minds: A comprehensive character education framework* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Harrington, N. G., Giles, S. M., Hoyle, R. M., Feeney, G. J., & Yungbluth, S. C. (2001). Evaluation of the All Stars character education and problem behavior prevention program: Effects on mediator and outcome variables for middle school students. *Health Education and Behavior, 28*, 533–546.

Lapan, R. T., Gysbers, N. C., & Sun, Y. (1997). The impact of more fully implemented guidance programs on the school experiences of high school students: A statewide evaluation study. *Journal of Counseling and Development, 75*, 292–302.

Leming, J. S. (2000) Tell me a story: An evaluation of a literature-based character education programme. *Journal of Moral Education*, *29*, 413–427.

Lickona, T. (1991). *Educating for character*. New York: Bantam Books.

Lickona, T. (1995). *School as a caring community profile*. Cornell University: Author.

McWhirter, J. J., McWhirter, B. T., McWhirter, E. H., & McWhirter, R. J. (2004). *At-risk youth: A comprehensive response* (3rd ed.) Belmont, CA: Brooks/Cole.

Schultz, L. H., Barr, D. J., & Selman, R. L. (2001). The value of a developmental approach to evaluating character development programmes: An outcome study of Facing History and Ourselves. *Journal of Moral Education, 30*, 3–27.

Williams, D. D., Yanchar, S. C., Jensen, L. C., & Lewis, C. (2003). Character education in a public high school: A multi-year inquiry into Unified Studies. *Journal of Moral Education, 32*, 3–33.

Williams, M. M. (2000). Models of character education: Perspectives and developmental issues. *Journal of Humanistic Counseling, Education and Development, 39*, 32–40.