

METHODOLOGICAL CONSIDERATIONS IN EVALUATING SCHOOL-BASED PROGRAMS TO PROMOTE SOCIAL COMPETENCE AND REDUCE PROBLEM BEHAVIOR

Greta M. Massetti

University at Buffalo, SUNY

Hugh Crean and Deborah Johnson

Children's Institute and the University of Rochester

David DuBois and Peter Ji

University of Illinois at Chicago

Interventions that aim to promote social competence, reduce problem behavior, and improve school climate are common at all levels of schooling. This whole-school focus, coupled with researchers' concerns about contamination or spillover effects in evaluations that randomly assign classrooms or students to conditions, as well as advances in statistical modeling, have resulted in an increased emphasis on the use of cluster-randomized trials in education settings. That is, evaluation of educational interventions often entails random assignment of schools, rather than classrooms or individual students, to intervention and treatment-as-usual groups. Such evaluations present unique methodological challenges with respect to the recruitment of schools and the choice of methods to ensure equivalence of groups. Furthermore, active consent return rates may be significantly challenged in school-based research, particularly in urban, high-risk schools. In this article, we examine issues of school recruitment, equivalence of intervention groups, and consent rates in a sample of 42 schools from 3 sites that participated in the Social and Character Development Research program funded by the Institute of Education Sciences (IES), in collaboration with the Centers for Disease Control and Prevention (CDC). Although differences between sites are apparent, the intervention and comparison groups of schools within each site are shown to be equivalent on both a range of school demographic characteristics and on consent return rates. Implications for conducting randomized evaluations of school-based intervention and prevention programs are discussed.

• **Greta M. Massetti**, 4770 Buford Highway NE, MS F-64, Atlanta, GA 30341. Phone 770.488.3943.
E-mail GMassetti@cdc.gov

Journal of Research in Character Education, 7(2), 2009, pp. 103–117
Copyright © 2009 Information Age Publishing, Inc.

ISSN 1543-1223
All rights of reproduction in any form reserved.

In accordance with the view that strengthening the climate of the whole school is critical for promoting social development and academic achievement and preventing problem behaviors (Greenberg, Domitrovich, & Bumbarger, 2001), programs and interventions to promote social competence and reduce problem behavior are commonly implemented throughout schools, at all grade levels (Gottfredson & Gottfredson, 2001). Likewise, contamination or “diffusion of treatment” effects in evaluations that randomly assign classrooms or students to conditions (Cook & Campbell, 1979; Flay & Collins, 2005) may lead participants randomly assigned to one condition to influence participants not assigned to the condition, such as teachers trained in a novel instructional strategy sharing the training with their colleagues at the same school. This concern with contamination effects, coupled with the aforementioned focus on school climate and contextual influences on intervention impact, have resulted in increased emphasis on the use of cluster-randomized trials (CRTs) in evaluation of educational interventions in which random assignment to intervention and treatment-as-usual groups is at the level of schools, rather than classrooms or individual students. Resource and logistical challenges are common among such evaluations given the intensive efforts and collaborations required to adequately assess intervention impact. Because of these challenges, school-based evaluations are often conducted with a relatively low number of schools. This reality, coupled with the potential for large variations in school characteristics and implementation strategies, can raise concern about baseline treatment-control group equivalence and statistical precision in estimating program effects.

Concerns with statistical power and statistical precision have been addressed in depth by recent work on optimal design for CRTs (Bloom, 2005; Raudenbush, 1997; Raudenbush, Martinez, & Spybrook, 2007). Recent research has made some progress in using CRT designs creatively to address ethical and practical concerns with randomization (Borman,

Slavin, Cheung, Chamberlain, Madden, & Chambers, 2005). Within the CRT design, the predominant influence on statistical power is the number of clusters, rather than the number of participants per cluster (Blair & Higgins, 1985; Bloom, 2005). In these considerations, investigators must address the relative cost of sampling more clusters as well as more participants per cluster.¹ For example, in a school-based intervention study, recruiting each additional school and ensuring that school’s continued commitment to the intervention throughout the duration of the study is far more expensive than sampling a greater number of students within each school. In order to maximize statistical power in cluster-randomized designs, efficient experimental designs and analytical approaches must be used. One common approach involves blocking prior to randomization (Raudenbush et al., 2007). Blocking involves matching clusters (in this case, schools) on relevant variables and conducting random assignment within blocks, such as school size or percent of students enrolled in the free or reduced lunch program. Blocking can maximize statistical power to detect significant program impacts in group-randomized designs.

The emphasis in evaluation using CRT designs is on addressing both the child-level intervention context, such as information about the curriculum, the skills it teaches, and the targeted outcomes, as well as the teacher, classroom, and school-level contexts. For example, if one of the main activities in the intervention involves training teachers to deliver a curriculum or technique, the program is likely to impact not only the students’ behavior and skills but also those of teachers, such as improvement in teacher knowledge, attitudes, or confidence about their own teaching effectiveness. Likewise, school-level variables can be impacted, such as school disciplinary practices as a result of participation in a unified intervention approach. This emphasis on implementing interventions and evaluating impacts at multiple levels of the school ecology (e.g., student, classroom, and

school) is particularly relevant for group-randomized designs, where intervention impacts may extend beyond the level of the individual student. For example, Cook et al. (1999) reported the impact of a whole-school reform approach on both individual student outcomes as well as the schools' social climate, teachers' self-efficacy, and students' social behavior. Thus, the use of these designs allows researchers to evaluate the impact of an intervention on *place*, such as classrooms or schools, as well as the impact on *individuals*.

The implementation of school-based CRT evaluations requires attention to a challenging mix of methodological, ethical, and pragmatic issues. How investigators address these concerns influences the degree to which researchers and practitioners can conclude that differences between intervention and treatment-as-usual groups are due to the intervention itself rather than to other factors (internal validity) and that program effects can generalize to other schools (external validity). Included among the methodological challenges in conducting school-based group-randomized designs and that are considered in this paper are the following: (1) the extent to which schools in the study are representative of the larger population of schools to which the findings can be generalized; (2) the extent to which intervention and comparison groups are equivalent on baseline characteristics of students and schools; and (3) validity of the evaluation process, including consent procedures and participation/survey return rates.

Recruitment issues present unique challenges in a school-randomized design, as there is typically a limited sample of schools that are eligible for the study (Drews et al., 2009; Ji, DuBois, Flay, & Brechling, 2008; Rice, Bunker, Kang, Howell, & Weaver, 2007). Furthermore, concerns with the potential of being assigned to the comparison or control group may cause resistance or reluctance on the part of administrators and decision makers (Ji et al., 2008; Kam, Greenberg, & Walls, 2003). A lack of understanding of random assignment procedures and the value of randomized trials

contributes to confusion and skepticism on the part of teachers and school administrators. However, if schools participating in CRTs of school-wide interventions do not have characteristics or student populations that are representative of the intended population of schools or students to which the evaluation is intended to generalize, the external validity of the study is threatened. This is especially true of studies examining the impact of programs that intend to target high-risk students and schools. If findings can not be generalized to the sample of students and schools that are most in need of the intervention or program that is being evaluated, the policy impact of the study may be severely limited. Given that the recruitment of schools that are willing to participate in a school-randomized trial necessarily places some selection bias on the sample of eligible schools, it is important to compare the extent to which participating schools are similar to the broader population of schools from which they are sampled on as many relevant characteristics as possible.

Following recruitment of schools for school-based randomized evaluations, procedures are typically required for securing active informed consent from parents of participating students. In such studies, informed consent to participate in data collection to evaluate program impacts must be collected from primary caregivers of students eligible to participate in the study. However, if the programs implemented in school-based evaluations constitute "standard educational practice," researchers are not required to collect informed consent for students to participate in the program or intervention itself. Active informed consent refers to the requirement that primary caregivers must sign a document providing permission for their child to participate and return the document to researchers (as opposed to only returning the document if they do *not* want their child to participate, referred to as passive consent). Informed consent documents typically have check boxes on the form for primary caregivers to indicate whether they do or do not give permission for their child to partici-

pate in data collection. If primary caregivers do not return a signed consent document, it is assumed that they do not wish their child to participate in data collection.

Securing active consent from parents is an important part of school-based research, as consent forms allow parents to understand the goals of the study and to make informed decisions about their children's research participation. These processes ensure that human subjects are protected and the research protocol conforms to ethical guidelines. However, securing active consent presents some practical concerns for researchers. Studies have documented problems with active consent procedures in school settings due to such factors as limited contact with research staff and low investment of parents and students in the research process. Accordingly, low consent return rates have been documented as a common problem in school-based research (Bergstrom et al., 2009; Pokorny, Jason, Schoeny, Townsend, & Curie, 2001; Rice et al., 2007; Stein et al., 2007), particularly when active consent is required. Furthermore, several studies have demonstrated a sample bias in active consent procedures, in which parents who complete and return active consent documents are not representative of the broader population of parents in the classroom, with high-risk students being less likely to return consent forms (Esbensen & Deschenes, 1996; Frame & Strauss, 1987; Noll, Zeller, Vannatta, Bukowski, & Davies, 1997; Pokorny et al., 2001; Severson & Biglan, 1989). Efforts to increase consent returns often require substantial investments in time and resources on the part of study personnel (Ellickson & Hawes, 1989; Fletcher & Hunter, 2003). Researchers conducting school-randomized evaluations must be very attentive to ensuring the protection of human subjects while weighing the relative costs and benefits of investing staff time and resources in increasing consent rates.

The present study examines methodological considerations in conducting CRT evaluations of three school-wide social and character development programs as part of a multipro-

gram evaluation, the Social and Character Development (SACD) Research Program, funded by the Institute of Education Sciences (IES), in collaboration with the Centers for Disease Control and Prevention (CDC) (see Haegerich & Metz, this volume). The SACD Research Program involved the evaluation of the efficacy of seven elementary school-based and school-wide programs that intend to promote social competence, reduce problem behavior, and promote school climate. In the multisite SACD Research Program, seven groups of investigators recruited a sample of participating schools, which were matched within pairs and then randomly assigned within each pair to participate in the Intervention or Comparison (traditional educational practice) groups for the 3-year study. The target cohort of students for the study consisted of third graders in the first year of implementation of the study at each participating school and was followed for 3 years. Schools were recruited, matched, and randomly assigned to the Intervention and Comparison groups prior to year 1 of the study. During year 1, active informed consent was secured from parents of third grade students in the fall, just prior to data collection procedures.

This paper examined how data on school characteristics and consent rates can inform understanding of the internal and external validity of the evaluation. Data on program impacts will not be included presently, as the focus of this study was on methodological considerations and their implications for internal and external validity. A full report including program impacts for all participating sites will be included in a report authored by the SACD Consortium and released by IES. Information will be provided for three participating sites in the SACD Research Program: the University at Buffalo, SUNY (UB), the University of Illinois-Chicago (UIC), and the Children's Institute, in collaboration with the University of Rochester (CI). These three sites were included because they had collected comparable data regarding recruitment and consent procedures, so that information could be com-

bined and analyzed across sites. Information about school recruitment, consent procedures, and design will be provided for the three sites as will descriptive information for the sample of participating schools. Analyses explored three issues: (1) the extent to which schools in the study are representative of U.S. schools; (2) equivalence of intervention and comparison groups; and (3) consent participation for students. The analyses and subsequent results will seek to establish the degree to which findings from the present evaluations of programs can be generalized to other school populations, the degree to which program impact findings are valid, and whether intervention findings are reflective of the entire student body, not just those students who participated in data collection efforts.

METHOD

Participants

A total of 42 schools participated in the SACD Research Program at the UB, UIC, and CI sites, with 14 schools at each site matched and randomly assigned in equal numbers to the Intervention and Comparison groups. The schools were recruited by study personnel at each site following some initial discussions at the multi-program level regarding the goals and procedures for recruitment. Each site devised plans for recruiting and ensuring schools' commitment independently. Although many of the criteria for school recruitment and selection were different across sites, the commonality across sites was that schools eligible for recruitment needed to be open to and agree to random assignment and implementation of the selected program if assigned to the intervention group. Following recruitment and selection of participating schools, all sites used a matching program provided by Mathematica Policy Research, Inc. (2007), the independent multiprogram evaluator, to all investigators that used an algorithm to select the best pairs by minimizing the distance between several mea-

surable characteristics for schools within each pair. A set of candidate pairs was selected such that the overall quality of matches across all the schools remained as high as possible, without creating any serious mismatches for any subset of the individual pairs. Each site executed the matching algorithm with somewhat different school characteristics variables, depending on what data were available, and exercised its best judgment, based on knowledge of the schools involved, as to which candidate pairing was the best one. After the best pairing was established, one member of each pair was randomly selected to be in the treatment group and the other was assigned to the control group.

School Recruitment Procedures

University at Buffalo, SUNY. Schools were recruited by the University at Buffalo (UB) to participate in a randomized trial of the Academic and Behavioral Competencies Program (Pelham et al., 2005; Waschbusch, Pelham, Massetti, & Northern Partners in Action for Children and Youth, 2005). School recruitment at UB began with the 50 elementary schools in one participating school district. Of these 50 schools, district personnel identified a list of 18 schools that would be eligible to participate based on grades (kindergarten through Grade 5), minimum enrollment (50 students per grade), district eligibility (not scheduled for closure or redistricting in the 3 years of the study), and demographics (more than 50% of students below poverty). Principals at these 18 schools were invited to a meeting with study personnel where the study was described. Principals who were interested in participating scheduled presentations to teachers and school staff by study personnel; the presentations were followed by a vote of all school teachers and staff members. Ten out of 13 schools elected to participate in the study. After letters of commitment were obtained by all ten principals, one of the principals strongly indicated that he would withdraw his school if it were randomized to the comparison group. Due to matching procedures, if one school dropped

out of the study following randomization procedures the matched school in the pair would also have to be excluded from the study. Therefore, efforts were made to secure the principal's commitment prior to matching and random assignment. Following multiple conversations with the principal and with district officials, it was determined that the risk was too great to lose both schools in the pair, and that school was dropped, providing a sample of nine schools. Due to matching procedures, it was imperative to generate pairs of schools that were closely matched, requiring an even number of schools. As the sample of schools in the original district was exhausted, two local charter schools and two suburban districts that were in close geographic proximity to the selected sample of schools were contacted to recruit additional schools. UB investigators contacted both charter and suburban schools to maximize chances that two schools with similar demographic characteristics would agree to participate. Two charter schools and two suburban schools agreed to participate in the study, securing a sample of thirteen schools (two charter schools, two suburban schools, and nine urban schools from the original district).

Demographic characteristics for 13 schools (9 in the city school district, 2 charter schools, and 2 suburban schools) were entered into a matching program provided by the contractor for the SACD study (Mathematica Policy Research, 2007), including: district, enrollment, percent enrollment in free or reduced lunch program, percent minority enrollment, and performance on state-administered achievement tests. The program yielded 6 pairs of schools, with one additional school that did not provide a good match with any of the others. All 12 schools that provided an adequate match were retained for the study. These 12 schools were randomly assigned within pairs to the Intervention and Comparison conditions.

In the first year of the study, IES staff approached study personnel and provided an option for additional funding for sites to

increase the number of schools to address concerns about the study's power to detect program impact. Two additional schools in the city school district were recruited and randomly assigned to the Intervention and Comparison groups. These two schools had not been in the original sample of schools recruited for the study, due to the fact that they had not met criteria for inclusion. One school was new, resulting from a merging of one school including grades pre-K to 4, and another school with Grades 5 through 8. The second school was a small school that had in the previous year increased enrollment from fewer than 25 third grade students to just under 50 third grade students, thereby making the school eligible to participate in the study. Information from these schools was entered in the matching program, and it was determined that the schools were an adequate match for each other, and for the original sample of schools. These two additional schools were randomly assigned to the Intervention and Comparison groups, and became part of the second cohort of schools in the SACD Research Program. This second cohort began participating in the study in the second year of implementation.

University of Illinois at Chicago. Schools were recruited by the University of Illinois at Chicago (UIC) to participate in a randomized trial of the Positive Action program (see Flay, et al., this volume). The participating schools for the UIC site were drawn from an initial pool of 483 elementary schools in the Chicago Public Schools (CPS). Schools were *excluded* from this pool if (1) they were not community schools (i.e., were academy, charter, special education), (2) they were already using the Positive Action program (the program to be implemented in the intervention schools) or similar intervention programs, (3) their enrollment rate was below 50 or above 140 students per grade, (4) their annual student mobility rates were 40% or above, (5) more than 50% of their students passed the Illinois State Achievement Test (ISAT), or (6) less than 50% of their students received free lunches.

Using these criteria, 68 schools were eligible to participate.

Following informational sessions at schools conducted by study investigators, 18 schools agreed to participate. School demographic variables were entered into the matching program referred to previously (Mathematica Policy Research, 2007). Matching variables used at the UIC site included total enrollment, percent minority enrollment, performance on state-administered achievement tests, free or reduced lunch eligibility, attendance, truancy, mobility, parent involvement, nonqualified teachers, and neighborhood crime. Of the nine pairs generated by the matching program, seven were selected based on best match. The schools in each pair were then randomized to the Comparison and Intervention conditions. Notification regarding group assignment was done through in-person visits by study personnel with school administrators. All schools in the seven pairs were successfully recruited and retained in the study (Ji et al., 2008). A detailed agreement was signed by both the principal and the president of the Local School Council.

Children's Institute. Schools were recruited by the Children's Institute (CI) to participate in a randomized trial of the Promoting Alternative Thinking Strategies program (see Flay et al., this volume). School recruitment at CI began with a sample of 10 schools in Minnesota and New York that had been involved in a prior collaboration with the University of Rochester and CI personnel. Although all 10 schools initially agreed to participate in the study at the time of the funding application, four schools subsequently declined participation due to circumstances outside the researchers' control (such as ongoing involvement in other studies that precluded participation in the SACD Research Program or initial misunderstandings regarding the implications of randomization). Study personnel then recruited an additional four schools to fill the sample of ten schools. School recruitment procedures at CI involved initial selection based on schools already implementing the Primary Mental Health Project (Cowen & Hightower, 1996), a

school-based prevention program for selected at-risk kindergarten through third grade children. Principals at schools implementing Primary Project were contacted and invited to participate in the study. Those principals who expressed interest then determined buy-in from teachers and parents before committing to participate in the study. Matching variables for the CI site included: location/district; total enrollment; percent minority enrollment; percent enrollment of English Language Learners; student-teacher ratio; percent eligible for free or reduced-price lunch; student mobility; percent at or above grade-level mastery on the state-administered tests in English Language Arts and Mathematics. Schools in the five matched pairs were then randomly assigned to the Intervention and Comparison groups. Randomization within the matched pairs occurred via coin flip conducted by a CI researcher not involved with the project.

As with UB, during the first year of implementation of the study, personnel were approached by IES and invited to apply for additional funding to recruit an additional 4 schools to increase the sample to 14 schools. The same school recruitment procedures (initial principal contact, recruitment of teacher buy-in) were followed for this second cohort of schools, which were then paired and randomly assigned. The second cohort of schools began implementation of the program in the second year of the study. The final sample of participating schools tended to be more heterogeneous, with 8 urban schools (4 cohort 1 and 4 cohort 2) and 6 suburban schools participating.

Consent Procedures

University at Buffalo, SUNY. Consent forms for primary caregivers to provide permissions for students to participate in data collection procedures were distributed to all third grade students in September of the first year of the study by research staff. Multiple levels of incentives were offered for return of consent forms. First, all classes that returned at least 85% of the signed consent forms, regardless of

whether they were negative or positive consent, received a pizza party. Furthermore, each teacher received a \$25 gift card for a store that sold educational materials for 85% return rates. If all third grade classes returned 85% or more of the consent forms, a \$250 donation was made to the school for the use of the third grade students. These were typically used for field trips and special events.

One staff member was assigned to each participating school to collect consent forms. Each staff member visited the classrooms a minimum of two times a week from the second week of school until the week prior to data collection (a period of approximately 5 weeks). During each visit, the staff member brought small gift bags with colored pencils, stickers, and other small prizes to give out to each student who returned a consent form. Additionally, a consent return report was given to teachers, and a presentation was made to students to remind them of their progress in collecting consents and towards earning the pizza party. Across all third graders in the sample, 90.5% of students returned signed consent forms. Of these, 71.1% were positive consents.

University of Illinois at Chicago. Parent consent was obtained in September of the first year of the study. Research staff visited each classroom and distributed the consent form for students to take home to their parents. As an incentive, a pizza party was offered for all students in a classroom if consent forms were returned by 90% or more of the students. Research staff visited each classroom for up to 4 consecutive days to collect consent forms. A large visual “thermometer” that showed the percentage of forms returned was displayed in each classroom. Teachers were instrumental in assisting the research team with collection of consent forms and thus were offered a gift certificate if a return rate of 90% was achieved for all forms (both positive and negative consent). Out of all third grade students in the participating schools, 98.3% returned signed consent forms, 79.7% of which were positive consents.

Children’s Institute. Parents of potential third grade students in each of the participating

schools were provided a letter from the principal explaining the research program and why the school decided to participate, accompanied with the parental consent documents. The first round of consent documents was either mailed to parents or given to students at school in May prior to year 1 of the study for four schools and in September of year 1 for the remaining schools. A pizza party was offered to classes that returned 85% of their consent forms (similar to UIC, a large pizza “thermometer” that displayed the percentage of forms returned was displayed in each classroom). Individual students who returned a consent form were given a small prize or trinket (e.g., pencil, refrigerator magnet, etc.). Teachers were contacted each week to update consent progress and revisit strategies, noting the children who had not returned consent forms. For the school that began consent procedures in the spring, consent procedures were continued in the fall of year 1. Additionally, participation by CI research staff at open house events for parents and parent liaisons were used to increase the rate of consent returns. Across all third grade students in the participating schools, 84% returned signed consent forms (88% of the suburban students; 78% of the urban students). Of these, 70.5% were positive consents (72.9% of the urban students; 66.3% of the suburban students).

Design and Data Collection Procedures

Demographic characteristics of participating schools from the three sites (see Table 1) were gathered through publicly available data sets (such as the Common Core of Data, Sable & Hill, 2006); or school district reports) on a range of school characteristics, including school enrollment, grade structure, poverty rates, minority enrollment, and academic achievement, for the 2004–2005 school year (Year 1 baseline). Total school enrollment was defined as all students enrolled in the same building. For the UB site, 12 of the 14 schools included grades pre-kindergarten through 8. The other two schools included grade kinder-

TABLE 1
Descriptive Characteristics of Participating Intervention and Comparison Schools

	Total Mean (SD)	Intervention Mean (SD)	Comparison Mean (SD)	F p
Total enrollment	548.8 (179.9)	513.4 (140.7)	584.1 (209.5)	1.65 .21
Free or Reduced Lunch*	77.9 (29.1)	78.3 (27.3)	77.5 (31.5)	.01 .93
Percent minority	72.1 (29.9)	72.9 (90.9)	71.2 (29.6)	.04 .85
Number of grade levels	8.12 (1.55)	8.05 (1.60)	8.19 (1.54)	.09 .77
Combined achievement**	59.4 (17.9)	58.6 (17.2)	60.3 (18.9)	.09 .76
Percent consented***	70.7 (11.9)	71.9 (9.8)	69.5 (14.0)	.43 .51

N = 42 schools, 21 Intervention and 21 Comparison

* Free or Reduced Lunch = Percent of students enrolled who are eligible for Free or Reduced Lunch Program.

** Combined Achievement = Mean of percent of students scoring at or above grade level on Grade 4 Mathematics and Grade 4 English Language Arts state-administered tests.

*** Percent consented = Percent of Grade 3 students who returned a signed consent for permission to participate in the study prior to baseline data collection.

garten through 5. For the UIC site, 9 schools had kindergarten through Grade 8; four had kindergarten through Grade 6; and one had kindergarten through Grade 5. For the CI site, 4 schools had kindergarten through Grade 5, and 10 schools had kindergarten through Grade 6. Because the grades were considered enrolled in the same schools and were overseen through a single administrative body, the enrollment for all grades was included for all schools. For the poverty rates, percent of students eligible for free or reduced price lunch was used in the demographic characteristics of schools. Percent minority enrollment was a measure of student diversity (all non-White, non-Hispanic students were counted as minority). For academic achievement each school's percent of students performing at or above grade level on the state-administered achievement tests in Mathematics and English Language Arts in Grade 4 during the first year of implementation of the study were averaged. These variables were chosen because they are likely to reflect important differences across schools and because they were available from all sites and measured consistently across sites.

RESULTS

Differences in School Characteristics Between Intervention and Comparison Conditions, and Across Sites

Four two-way ANOVAs (Intervention vs. Comparison as one factor and site as the other factor) were conducted to examine potential differences between Intervention and Comparison schools on total enrollment, percent of students eligible in free or reduced lunch programs, percent minority enrollment, and school academic achievement. The F-values are presented in Table 1, along with means and standard deviations for both groups, as well as the overall mean for the sample of schools. For total school enrollment, there were no main effects of intervention group or site, and no interaction ($ps > .20$). For percent eligible for free or reduced price lunch, there was no main effect of intervention group ($p > .20$), but there was a significant site effect, $F(2, 36) = 4.09$, $p = .03$. Post-hoc tests indicated that the CI site had a significantly lower percentage of children

eligible for free or reduced lunch (61.88%) compared to the UIC site (91.98%); the UB site (79.84%) was not significantly different from either. There was also a main effect of site on percent minority enrollment $F(2, 36) = 5.21, p = .01$, with the CI site (59.04%) and the UB site (65.64%) having lower minority enrollment than the UIC site (91.46%). There was no significant effect for intervention group on minority enrollment ($p > .20$). For school academic achievement, there was no main effect of intervention group ($p > .20$), and there was a significant effect of site, $F(2, 36) = 8.56, p = .001$, with the UIC site (46.81) having lower average performance on state tests than both the UB (60.25) and the CI (71.21) sites.

Differences in School Characteristics Between Study Schools and the U.S. School Population

School characteristics of the participating schools were compared to the overall average of U.S. schools available through the Common Core of Data (Sable & Hill, 2006) using one sample t-tests, with the national average as the null value. Participating schools in the study were larger (total enrollment of 548.8 compared to 445.2 for U.S. schools, $t(41) = 3.73, p < .01$), had higher rates of poverty (77.9 eligible for free or reduced lunch compared to 41.6 for U.S. schools, $t(41) = 8.08, p < .001$), and had higher enrollment of minority students (72.1% compared to 41% for U.S. schools, $t(41) = 6.73, p < .001$). These differences are indicative of the fact that the majority of schools that participated in the study from the UB, UIC, and CI sites were in urban settings.

Differences in Consent Form Return Rates Across Sites

A two-way ANOVA was conducted to compare the Intervention and Comparison groups on the percentage of students in the third grade cohort who returned positive parental consent to participate in the study

prior to fall data collection in the first year of the study. For Intervention schools, 71.9% of students provided signed consent forms, whereas 69.5% of students in Comparison schools provided informed consent. This difference was very modest and not statistically significant, indicating that the groups did not differ in participation in the fall wave of data collection in year 1. The main effect of site was significant, $F(2, 35) = 10.93, p < .001$; UIC's consent rates (80.64%) were significantly higher than those of UB (66.73%) and CI (64.31%).

DISCUSSION

The present study provides information on the recruitment and consent collection procedures for 3 sites participating in the SACD Research Program funded by IES, in collaboration with CDC: the University at Buffalo, SUNY, the University of Illinois at Chicago, and the Children's Institute. The 42 schools were recruited to participate in a matched, school-randomized study to evaluate the impact of universal (delivered to all students in schools), school-wide programs that all aimed to promote social development, reduce problem behavior, and improve school climate. Schools participated in the study for 3 years, and the study followed students who were enrolled as third graders in the first year of implementation of the study at each participating school. The present study examined the comparability of the Intervention and Comparison schools at the three sites on a range of school characteristics, including total enrollment, poverty (as indicated by eligibility for free or reduced price lunch), minority enrollment, and performance on state-administered standardized tests. Findings indicate that Intervention and Comparison schools were similar at the beginning of the study on all variables and measured characteristics. Thus, there can be greater confidence that any differences found between Intervention and Comparison schools after 1, 2, and 3 years of program implementation are

due to program impacts, rather than to preexisting differences between groups at baseline.

It should be noted, however, that Intervention and Comparison schools may have differed on other, unmeasured characteristics. These include some factors, such as a school's readiness to implement a universal, school-wide intervention, that have the potential to significantly influence student outcomes and thus may bias intervention impact estimates (Ji et al., 2008). A similar caution applies to the degree to which participating schools might be assessed to be similar to nonparticipating schools at each site. This latter consideration underscores the need for further investigation into procedures that increase the proportion of schools that demonstrate a willingness to participate in evaluations of school-wide interventions employing random assignment, which overall was quite low (26%) in the present study for at least one of the sites (UIC). Such procedures might involve establishing long-term partnerships between researchers and school districts to ensure the mutual understanding and investment in the research enterprise by all parties, for example.

Significant effects of sites were found for three school characteristics: percent of students eligible for free or reduced-price lunch, percent minority enrollment, and school academic achievement. Effects likely reflect the greater homogeneity and urbanicity of the UIC schools, as the CI and UB sites included a mix of urban and suburban schools. While it is not possible to evaluate the impact of site effects on program outcomes, as each site implemented and evaluated a different program, and programs thus vary across sites, site differences should be taken into consideration in analyses evaluating program effects.

Furthermore, the urban status of the majority of participating schools at the three sites is reflected in the average size of the schools and in the higher rates of poverty and minority enrollment compared to average of all U.S. elementary schools available on the Common Core of Data. The participating schools at the

UB, UIC, and CI sites represented samples of predominately large, urban, high-risk schools, with greater variability along this dimension for schools at the UB and CI sites. Given the need for effective programs to target social and character development, minimize disruptive and negative student behavior, and promote student competencies in such high-risk schools, the present sample of schools provide an important context for evaluating the impact of the interventions.

Concerns with recruitment of students in data collection efforts, particularly in high-risk settings (Dent et al., 1993; Esbensen & Deschenes, 1996; Pokorny et al., 2001; Severson & Biglan, 1989), suggest that efforts to maximize returns of active consent forms must be conducted in ways that are mindful of the need to ensure protection of human subjects. The participating sites in the present study received active participant consent forms from 71% of the sample, which is a higher rate of consent compared to past research on school-based implementation of similar research (Esbensen & Deschenes, 1996; Frame & Strauss, 1987; Noll et al., 1997). Given the higher proportions of students living in conditions of poverty for the participating schools, the rates of active consent returned for these schools indicate successful consent recruitment efforts, and indicate that program impact findings can be attributed to true program effects, rather than to the programs being tested with a biased sample. This is likely due to the extensive investment in time and resources that each site made in distributing and collecting consent forms to ensure high participation of third graders in the data collection procedures. For example, all sites participated in regular visits to schools by project staff to collect consent forms and encourage participation, thereby increasing the visibility of the study to the schools' staff and to the students. There was a significant effect of site for the rate of returns of active consent forms, indicating that the UIC site had higher rates of consent than did UB and CI. The consent procedures and incentives were very consistent

across sites. For example, all sites employed a combination of teacher and school staff engagement, pizza parties and other incentives, and regular contact with teachers and students to maintain motivation and enthusiasm for consent returns. Although it is not possible to determine empirically to what the site differences can be attributed, it was the impression of the investigators at the UIC site that there is a high expectation for return of forms for students at the schools in general. Furthermore, the schools that participated in the UIC site had weekly assignment folders that were used to send home consent forms and minimized risk that forms would be lost. This process might have benefited the consent process at those schools and may have accounted for the significant difference in return rates between sites.

An issue that tends to receive scant attention in the empirical literature but likely has implications for impact analyses is how schools get randomized. More specifically, how are schools recruited for participation in a place-based randomized study? In the SACD Research Program, at least two differing strategies were used in recruiting participating schools. The first approach, adopted by UIC, was to identify a larger number of potential matched pairs of schools than was necessary for the study and then approach the best matches for participation. The other approach, used by the UB and CI sites, was to obtain commitment from the exact number of schools needed for the study and then run the matching algorithm on only those schools. This differing strategy is to some degree reflective of a smaller number of schools available for participation in the smaller geographic regions that participated in the UB and CI sites' study. Essentially, there was not a large enough "pool" of schools at the UB and CI sites to recruit a large number prior to matching, and then narrow the list based on matching results. This is likely to be the reality in many school-based randomized evaluations, as all but those using the largest urban school districts are likely to have a relatively narrow pool

of schools from which to choose. Though the approach adopted by UIC may have led to better school demographic matches, the approach adopted by other sites may have resulted in better matches on other unmeasured characteristics (e.g., principal support for SACD). Differences in school recruitment procedures may have also been influential in other unmeasured ways given that detailed data (e.g., student or teacher surveys) were not collected from nonparticipating schools, thus making it impossible to fully determine the extent to which the participating schools constitute a unique sample of schools from their home districts. It is possible that the approach used by the UB and CI sites, which involved recruiting only the specific number of schools needed for the study, yielded unique samples of schools that had the particular willingness and capacity to implement SACD programs and participate in the research project. The UIC approach, on the other hand, may have resulted in a sample that was more representative of the overall pool of available schools, an advantage from the perspective of external validity. These possibilities underscore the need to carefully consider the implications of different recruitment and randomization strategies when designing school-randomized trials and to provide details regarding recruitment and randomization procedures so that findings can be interpreted in proper context.

Bloom (2006) noted that greater similarity within blocks and greater differences across blocks within cluster-randomized trials maximizes predictive power; the stratification approach utilized in the UIC recruitment strategies illustrates an effort to capitalize on these advantages at least with respect to increasing within block (in this case, within pair) similarity. One might suspect that the SACD Research Program would be in an ideal position to explore the issues of variability across block (pairs of schools) and potential implications for randomization and impact on analyses, given the large number of schools in the study. This is not the case, however, as some variables used for matching differed across

sites and were often measured in different ways. These issues do not threaten the validity of the randomization procedures that occurred within sites, as the most critical issue was to ensure that pairs of schools were well-matched within each site, rather than across sites. The differences in matching variables and procedures that were necessitated across sites, however, do raise important methodological questions about that merit careful consideration in the interpretation of findings from this and other similarly designed evaluations of school-level interventions.

Although issues of implementation fidelity are beyond the scope of the present paper, one important methodological issue in randomized evaluations of school programs involves the threats to internal validity that may be posed by having comparison schools that implement the intervention program. Whereas comparison schools at all three sites were offered the opportunity to receive training in the intervention after the end of the study, study assessment procedures were put in place to be able to gauge the extent to which similar intervention activities were taking place in comparison schools during the study period that could undermine the integrity of the randomized design. For example, all schools participated in annual principal interviews and teacher surveys that assessed the implementation of any Social and Character Development-type program or strategy in the school. These data indicate that although schools engaged in a variety of strategies to address social and character development, none of the comparison schools implemented the specific programs or strategies evaluated as part of the SADC Research Program (see Bickman et al., this volume, for details).

In this article we have taken advantage of a rich set of descriptive methodological information from a multisite evaluation of school-based character and social development interventions to illustrate different methodological strategies that may be used in evaluations of school-based intervention and prevention programs that employ randomized

assignment at the level of the school and their potential implications for the internal and external validity of. Such detailed information is often not reported or critically considered in the reporting of cluster-randomized trial evaluations of school-based interventions. Yet, as our analysis demonstrates doing has the potential to be valuable to both researchers and practitioners as they seek to interpret and synthesize findings from such studies effectively and for researchers as they attempt to build on prior methodological learnings in the design of their own trials. We hope that our present effort along these lines will encourage researchers to collect and carefully incorporate reporting of the types of information we have discussed in their future work.

AUTHOR NOTE

The findings reported here are based on research conducted as part of the Social and Character Development (SADC) Research Program funded by the Institute of Education Sciences (IES), U.S. Department of Education, under contract ED-01-CO-0039/0006 to Mathematica Policy Research (MPR), Princeton, NJ, in collaboration with the Centers for Disease Control and Prevention's Division of Violence Prevention (DVP), and the recipients of SADC cooperative agreements. The SADC Consortium consists of representatives from IES, DVP, and the national evaluation contractor (MPR), and each cooperative agreement site participating in the evaluation. Research institutions in the SADC program (and principal researchers) include: *IES* Amy Silverman, Edward Metz, Elizabeth Albro, Caroline Ebanks; *DVP* Tamara M. Haegerich (previously, IES), Corinne David-Ferdon, LeRoy Reese (Moorehouse School of Medicine; previously DVP); *MPR* Karen Needels, John A. Burghardt, Heather Koball, Laura M. Kalb, Peter Z. Schochet, Victor Battistich (University of Missouri—St. Louis); *Children's Institute* Deborah B. Johnson, Hugh F. Crean; *New York University* J. Lawrence Aber, Stephanie

M. Jones (Harvard University), Joshua L. Brown (Fordham University); *University at Buffalo, The State University of New York* William Pelham (Florida International University); Greta M. Massetti (CDC), Daniel A. Waschbusch (Florida International University); *University of Illinois at Chicago/Oregon State University* 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 of Maryland* Gary D. Gottfredson, Elise T. Pas, Allison Nebbergall; *University of North Carolina at Chapel Hill* Mark W. Fraser, Thomas W. Farmer (Penn State University), Maeda J. Galinsky, Kimberly Dadisman; and *Vanderbilt University* Leonard Bickman, Catherine Smith.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Institute of Education Sciences, Centers for Disease Control and Prevention, Mathematica Policy Research, Inc., or every Consortium member, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Acknowledgment: The authors wish to thank the students, parents, teachers, and schools that participated in this study, without whom the research would not have been possible. Greta Massetti is now a Behavioral Scientist in the Division of Violence Prevention within the Centers for Disease Control and Prevention. Address all correspondence to Greta Massetti, 4770 Buford Highway NE, MS F 63, Atlanta, GA 30341; gmassetti@cdc.gov

NOTE

1. For a full description and discussion of considerations of sampling clusters and participants within clusters, see Raudenbush (1997).

REFERENCES

- Bergstrom, J. P., Partington, S., Murphy, M. K., Galvao, L., Fayram, E., & Cisler, R. A. (2009). Active consent in urban elementary schools: An examination of demographic differences in consent rates. *Evaluation Review*, *33*, 481–496.
- Blair, R. C., & Higgins, J. (1985). A comparison of the power of the paired samples rank transform statistic to that of Wilcoxon's signed ranks statistic. *Journal of Educational Statistics*, *10*, 368–383.
- Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 115–172). New York: Russel Sage Foundation.
- Bloom, H. S. (2006). *The core analytics of randomized experiments for social research*. New York: MDRC.
- Borman, G., Slavin, R. E., Cheung, A., Chamberlain, A., Madden, N. A., & Chambers, B. (2005). Success for all: First-year results from the National Randomized Field Trial. *Educational Evaluation and Policy Analysis*, *27*, 1–22.
- Cook, T., & Campbell, D. (1979). *Quasi-experimentation*. New York: Rand McNally.
- Cook, T., Habib, F., Phillips, M., Settersten, R., Shagle, S., & Degirmencioglu, S. (1999). Comer's school development program in Prince George's County, Maryland: A theory-based evaluation. *American Educational Research Journal*, *36*(3), 543–597.
- Cowen, E. L., & Hightower, A.D. (1996). Primary Mental Health Project: School-based preventive interventions for adjustment problems. In M. C. Roberts (Ed.), *Model program in school mental health* (pp. 63-74). Hillsdale, NJ: Lawrence Erlbaum.
- Dent, C. W., Galaif, J., Sussman, S., Stacy, A., Burtun, D., & Flay, B. R. (1993). Demographic, psychosocial and behavioral differences in samples of actively and passively consented adolescents. *Addictive Behavior*, *18*, 51–56.
- Drews, K. L., Harrell, J. S., Thompson, D., Maz-zuto, S. L., Ford, E. G., Carter, M., et al. (2009). Recruitment and retention strategies and methods in the HEALTHY study. *International Journal of Obesity*, *33*, S21–S28.
- Ellickson, P. L., & Hawes, J. A. (1989). An assessment of active versus passive methods for

- obtaining parental consent. *Evaluation Review*, 13, 45–55.
- Esbensen, F. A., & Deschenes, E. P. (1996). Active parental consent in school-based research: An examination of ethical and methodological issues. *Evaluation Review*, 20, 737–753.
- Flay, B. R., & Collins, L. M. (2005). Historical review of school-based randomized trials for evaluating problem behavior prevention programs. *Annals of the American Academy of Political Social Science*, 599, 115–146.
- Fletcher, A. C., & Hunter, A. G. (2003). Strategies for obtaining parental consent to participate in research. *Family Relations: Interdisciplinary Journal of Applied Family Studies*, 52, 216–221.
- Frame, C. L., & Strauss, C. C. (1987). Parental informed consent and sample bias in grade school children. *Journal of Social and Clinical Psychology*, 5, 227–236.
- Gottfredson, G. D., & Gottfredson, D. C. (2001). What schools do to prevent problem behavior and promote safe environments. *Journal of Educational and Psychological Consultation*, 12, 313–344.
- Greenberg, M. T., Domitrovich, C., & Bumbarger, B. (1999). *Preventing mental disorders in school-aged children: A review of the effectiveness of prevention programs*. Washington, DC: U.S. Dept. of Health and Human Services, Center for Mental Health Services.
- Ji, P., DuBois, D., Flay, B. R., & Brechling, V. (2008). “Congratulations, you have been randomized into the control group!(!)”: Issues to consider when recruiting schools for matched-pair randomized control trials of prevention programs. *Journal of School Health*, 78, 131–139.
- Kam, C. M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science*, 4, 55–63.
- Mathematica Policy Research, I. (2007). *SAS programs for matching school pairs*. Retrieved from <http://www.mathematica-mpr.com/>
- Noll, R. B., Zeller, M. H., Vannatta, K., Bukowski, W., & Davies, H. (1997). Potential bias in classroom research: Comparison of children with permission with those who did not receive permission to participate. *Journal of Clinical Child and Adolescent Psychology*, 26, 36–42.
- Pelham, W. E., Massetti, G. M., Willson, T., Kipp, H., Myers, D., Newman Standley, B. B., et al. (2005). Implementation of a comprehensive schoolwide behavioral intervention: The ABC Program. *Journal of Attention Disorders*, 9, 248–260.
- Pokorny, S. B., Jason, L. A., Schoeny, M. E., Townsend, S. M., & Curie, C. J. (2001). Do participation rates change when active consent procedures replace passive consent? *Evaluation Review*, 25, 567–580.
- Raudenbush, S. W. (1997). Statistical analysis and optimal design for group randomized trials. *Psychological Methods*, 2(2), 173–185.
- Raudenbush, S. W., Martinez, A., & Spybrook, J. (2007). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis*, 29(1), 5–29.
- Rice, M., Bunker, K. D., Kang, D. H., Howell, C. C., & Weaver, M. (2007). Accessing and recruiting children for research in schools. *Western Journal of Nursing Research*, 29, 501–514.
- Sable, J., & Hill, J. (2006). *Overview of Public Elementary and Secondary Students, Staff, and Schools, School Districts, and Revenues, and Expenditures: School Year 2004–05 and Fiscal Year 2004 (NCES 2007-309)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Severson, H., & Biglan, A. (1989). Rationale for the use of passive consent in smoking prevention research: Politics, policy, and pragmatics. *Preventive Medicine*, 18, 267–279.
- Stein, B. D., Jaycox, L. H., Langley, A., Kataoka, S. H., Wilkins, W. S., & Wong, M. (2007). Active parental consent for a school-based community violence screening: Comparing distribution methods. *Journal of School Health*, 77, 116–120.
- Waschbusch, D. A., Pelham, W. E., Massetti, G. M., & Northern Partners in Action for Children and Youth (2005). The Behavior Education Support and Treatment (BEST) school intervention program: Pilot project data examining schoolwide, targeted-school, and targeted-home approaches. *Journal of Attention Disorders*, 9, 313–322.