

**A National Evaluation of Safe Schools/Healthy Students: Outcomes and Influences**

James H. Derzon<sup>a,\*</sup>, Ping Yu<sup>a</sup>, Bruce Ellis<sup>a</sup>, Sharon Xiong<sup>a</sup>, Carmen Arroyo<sup>b</sup>, Gary Hill<sup>c</sup>,  
Julia Rollison<sup>c</sup>

<sup>a</sup> Battelle Memorial Institute, Centers for Public Health Research and Evaluation, 2111  
Wilson Boulevard, Suite 900, Arlington, VA 22211, USA

<sup>b</sup> United States Department of Health and Human Services, Substance Abuse and Mental  
Health Services Administration, Center for Mental Health Services, 1 Choke Cherry Road,  
#6-1108, Rockville, MD, 20857, USA

<sup>c</sup> MANILA Consulting Group, Inc., 1420 Beverly Road, Suite 220, McLean, VA 22101,  
USA

Corresponding author; Tel.: +1 703 248 1640; fax: +1 614 458 0552; E-mail address:

[derzonj@battelle.org](mailto:derzonj@battelle.org) (J. Derzon)

## **Abstract**

The Safe Schools/Healthy Students (SS/HS) Initiative has awarded over \$2 billion in grants to more than 350 school districts in partnership with local mental health, law enforcement, and juvenile justice agencies. To estimate the impact of grantee characteristics, grant operations, and program characteristics in reducing violence and substance use, promoting mental health, and enhancing school safety, logged odds ratios (LORs) were calculated contrasting Year 3 with Baseline performance from grantee-provided data on seven outcome measures. After comparing grantee performance across outcomes and outcomes across grantees, the LORs were entered as dependent variables in a series of meta-regressions in which grantee characteristics, grant operations, and program characteristics were tested after controlling for pre-grant characteristics. Findings indicate that the SS/HS Initiative significantly improved most outcomes, that within-grantee performance varied greatly by outcome, and that random-effects meta-regression appreciably decreased the variance available for modeling. The approach demonstrates that the SS/HS Initiative is effective and that locally collected performance data can be used to estimate grantee success in improving youth outcomes.

## **A National Evaluation of Safe Schools/Healthy Students: Outcomes and Influences**

### **1. Introduction**

In the wake of a series of school shootings in the 1997–1998 school year that left 12 dead and another 47 wounded, Congress in 1999 enacted the Safe Schools/Healthy Students (SS/HS) Initiative. Developed as a joint effort of the Department of Education, the Center for Mental Health Services (CMHS) of the Substance Abuse and Mental Health Services Administration (SAMHSA), and the Department of Justice, the SS/HS Initiative was developed to promote the safe school environments that are essential to promoting healthy development and academic success and ensuring that students and their families feel connected to their school and their community. Since 1999, the SS/HS Initiative has awarded over \$2 billion in grants to more than 350 school districts in partnership with their local mental health, law enforcement, and juvenile justice agencies. These collaborations have in turn led to the implementation of locally designed, comprehensive plans to produce safe, respectful, and drug-free school environments, while promoting vital social skills and healthy childhood development.

Even 11 years after establishment of SS/HS, the program remains relevant. In 2009, 8 percent of America's students reported being threatened or injured with a weapon, such as a gun, knife, or club, on school property (Robers, Zhang, & Truman, 2011). During that same year, 31 percent of students in grades 9–12 reported that they had been in a physical fight at least one time during the previous 12 months anywhere and 11 percent reported being in a fight on school property. Children and youth are more likely to be victimized in school than out of school; in 2008, students aged 12 to 18 were victims of about 1.2 million nonfatal crimes (theft plus nonfatal crimes at school, compared to about 1 million nonfatal crimes away from school (Robers et. al., 2011). In addition, the most recent data related to bullying indicated that in 2007,

about 32 percent of students reported having been bullied at school. School violence, classroom disruptions, and bullying are all associated with lower student achievement (Lannie & McCurdy, 2007). The primary goal of the SS/HS program is to help schools to develop and implement comprehensive programs that are effective at creating a safe school community where students feel supported and can focus on learning. Because the consumption of alcohol and the use of illegal substances on school property by students has been shown to be associated with school violence, disruptive classroom behaviors, and academic failure (Kodjo, Auinger, & Ryan, 2003; Komro, Williams, Forster, Perry, Farbakhsh, & Stigler, 2000), reducing substance use is another goal of the SS/HS Initiative..

A third goal of the SS/HS Initiative is enhancing mental, emotional, and behavioral health. Overall, studies have shown a strong association between exposure to violence and mental, emotional, and behavioral problems including increased risk for depression, post-traumatic stress, aggression, delinquent behavior, and the perpetration of violent acts, such as fighting and bullying (Flannery, Singer, Van Dulmen, Kretschmar, & Belliston, 2007). Research has also shown that children and youth with unmet mental health needs are at greatest risk for engaging in violent behavior in school (Tolan & Gorman-Smith, 2002). Enhanced access to mental health services helps schools both to meet the needs of students at risk for engaging in violent or disruptive behaviors and to provide help for students who become victims of violence at school.

Based on research that has suggested that prevention efforts are most effective when families, schools, community organizations, and health care and service systems work together to implement programs and activities to help students (Epstein, 1995; Weissberg, Kumpfer, & Seligman, 2003), the SS/HS Initiative seeks to accomplish these goals by funding school-community collaborations connecting family, schools, and communities. School-community partnerships bring together the varying capacities, approaches, and missions of diverse

organizations and individuals to identify some of the issues that contribute to antisocial behavior, provide a platform for achieving consensus on shared goals and approaches, and establish frameworks for action (Lasker & Weiss, 2003a). However, the effectiveness of cross-agency partnerships in addressing broad health and social problems has been unclear (Mitchell & Shortell, 2000). Partnerships have the potential to falter because they are relationship-based and resource-intensive, and they often require people to work together in ways to which they are not accustomed (Kreuter, Lezin, & Young, 2000; Mitchell & Shortell, 2000; Wandersman, Goodman, & Butterfoss, 1997).

Nonetheless, there is widespread belief that attaining common goals and sustaining collaboration expand the capacity of the partnering organizations to address multidimensional issues (Lasker & Weiss, 2003b). Federal, state, and private foundations have increasingly required agencies to collaborate in order to receive funding (Butterfoss, Goodman, & Wandersman, 1993). Research and theory suggest there is great potential for partnerships to maximize power through joint action and to minimize duplication of services (Lasker, Weiss, & Miller, 2001; Butterfoss, 2007). Harnessing the capacities of these partners to create what Putnam (2000) calls “social capital” may be essential to maintaining school environments that are safe and foster the well-being of students.

To test whether and how the SS/HS activities and school-community partnerships are associated with improvement in youth outcomes, the current study examines the impact of the SS/HS Initiative on the 59 grantees funded in 2005 and 2006 on seven school and youth outcome measures: past 30-day alcohol use, past 30-day tobacco use, past 30-day marijuana use, perceived violence, experienced violence, access to school-based mental health services, and access to community-based mental health services. The study examines first individual grantee performance in improving each of the seven youth outcomes for which they reported data to determine if some

grantees performed better than others in improving youth outcomes. Next the study examines the overall impact of the Initiative across grantees to determine if the Initiative had an overall positive impact on youth. Lastly, the study examines the correlates of grantee success in improving youth outcomes by assessing the contribution of coalition structure and relations (grant operations) and the comprehensiveness with which each grantee implemented SS/HS-funded programs and activities after controlling for pre-grant differences. The program theory model and measures used to describe each component of the model tested are discussed in Rollison, Hill, Yu, and Murray (this issue). Due to the limited degrees of freedom available for modeling and lack of bivariate relations observed, not all variables listed in the program theory model were tested in the meta-regression.

## **2. Methods**

Data on pre-grant conditions and grant operations were collected using Census data; review of grant applications; site visits during each cohort's first year of funding; and annual surveys, interviews, and the review of grantee performance reports. The Initiative was unique in that it allowed grantees flexibility in reporting student performance data collected using local instruments and protocols and, at the time these data were collected, in only requiring grantees to submit performance data on one measure representing each of the three goals of the Initiative (violence/safety, substance use, and access to mental health services). These data and measures are described briefly, below.

### *2.1 Outcome (dependent) variables*

The SS/HS national evaluation collected grantee findings for seven outcome measures (past 30-day alcohol, tobacco, and marijuana use; experienced and perceived violence; and access to school- and community-based mental health services). Each year, using a variety of measures and instruments, grantees provided cross-sectional data on the number of students and the number or percent of those students displaying each outcome.

Because different measures were used to collect outcome data, it was necessary to standardize findings to ensure comparability across grantees. Thus, frequency data were dichotomized and logged odds ratios (LOR, see Formula 1) were calculated estimating the standardized change in each outcome comparing each available year of data (e.g., Year 3 and Baseline, Year 3 and Year 1, Year 3 and Year 2; see Fleiss & Berlin [2009] for a discussion of using LOR for dichotomous data). All outcome data were positively coded so that LORs greater than zero indicate that the grantee reported an improvement in the outcome over time, and LORs less than zero represent worse performance over time.

$$ES_{LOR} = \log_e \left( \frac{FaFd}{FbFc} \right) \quad \text{Formula 1}$$

where  $Fa$  = the number displaying the outcome and  $Fb$  = the number of subjects not displaying the outcome at Time 1,  $Fc$  = the number displaying the outcome and  $Fd$  = the number of subjects not displaying the outcome at Time 2.

Because grantee samples varied in size, when combining evidence across grantees all LORs were weighted by their inverse variance (Hedges & Olkin, 1985). These weights were calculated as ( $W_{LOR}$ , see Formula 2).

$$W_{LOR} = \frac{1}{SE_{LOR}^2} = \frac{abcd}{ab(c+d)+cd(a+b)} \quad \text{Formula 2}$$

Since LORs may not be familiar to social scientists, it is convenient for the sake of discussion to transform the logged odds into the more familiar standardized mean difference score  $d$  using the formula provided by Hasselblad and Hedges (1995; see Formula 3).

$$d = \frac{LOR\sqrt{3}}{\pi} \quad \text{Formula 3}$$

where  $\pi$  is the mathematical constant (approximately 3.14159).

Data were requested at baseline and annually, but SS/HS grantees were required to provide data for only one outcome per goal at each wave of collection. Fortunately, many provided data on multiple outcomes, but it must be noted that the amount of data varied by outcome and by grantee site. To increase the number of outcomes available for analysis, outcomes most closely representing grantee performance over the four waves of data, Baseline to Year 3, were included in the analysis (Chen & Shao, 2000; Rancourt, Sarndal, & Lee, 1994). Because of the potential that these estimates may introduce bias in the models (e.g., if longer periods are associated with better performance estimates), in each model, the number of years between estimates was tested prior to entering other variables in the regression models described below.

## *2.2 Pre-grant conditions, grant operations, and program characteristics (independent) variables*

Data from the 57 of 59 SS/HS grantees funded in 2005 and 2006 who contributed outcome data (96.6%) were collected from a variety of sources and selected for inclusion in the meta-regression based on program theory, completeness, and their distributions. Independent variables can be broken into three categories of potential influence, with the first being pre-grant conditions, the second representing grant operations, while the third represents the results of those grant operations, or how those grant operations were translated into local program priorities and activities. These variables are described below.

*Pre-grant environment control variables* represent extant conditions that may be related to underlying violence/safety, substance use, and other outcomes in grantee communities or to their potential to exhibit change in violence/safety, substance use, and other outcomes over time. The four pre-grant environment variables considered for this analysis are history of current partnership, poverty, grant recipient structure, and funding per targeted capita. Data for the pre-

grant environment variables were extracted from grant applications and other archival sources, such as Census, and were only available for the Baseline year. Because of extreme skew, funding per targeted capita was logged for analyses.

*Grant operations independent variables* reflect grantee operations in implementing the Initiative. Eight grant operations variables were selected for testing based on their completeness and likely capacity to explain differences in grantee performance. The grant operations variables tested in the meta-regression are indicators of partnership composition (represented by number of other partners), partnership organization, partnership interaction, partners' participation in decision making, partnership contributions, partnership functioning, perceived importance of school resources, and school involvement. Data on number of partners and partnership organization were collected through annual group telephone interviews of the participating project directors, local evaluators, and required partners. Partnership interaction and partnership functioning data were collected through an annual Partnership Inventory that was submitted by required partners. A web-based Project-Level Survey of the 57 grantees provided data on partners' participation in decision making, while data on partnership contributions were collected in annual telephone interviews of the participating project directors. Finally, data on perceived importance of school resources and school involvement in decision making were collected through an annual web-based survey of the approximately 1,500 schools participating in the 2005 and 2006 SS/HS cohorts. School responses were averaged to describe the typical school response for each grantee.

Due to extreme skew, two grant operations variables (number of other partners and school involvement in decision making) were logged for analysis. Categorical measures, such as

partnership organization, were recoded as binary variables, with 1 indicating the grantee had the characteristic and 0 indicating the grantee did not.

*Program characteristics independent variables* are variables that are under the control of the grantees as they implement the Initiative. The annual web-based Project-Level Survey collects implementation status data on 38 possible grantee activities. The activities are not ranked relative to one another nor are priorities obtained; grantees are simply asked the implementation status of each of the 38 activities. Included in this domain are indicators of coordination and service integration and comprehensive programs and activities, taken both as a whole (29 items) and as subsets according to outcome focus (i.e., comprehensive programs and activities for early childhood development [four items], mental health [six items], and school relationship with the community [eight items], respectively). Items summarized by these component scales were identified based on the face validity of items, and scale scores were created by averaging responses to items within each scale.

### *2.3 Statistical procedures*

*Main effects.* To develop best estimates of average performance on each of the SS/HS outcome measures, extreme weights and values were recoded (Winsorized) at two standard deviations above or below the mean of each distribution (Lipsey & Wilson, 2001). The grand mean for each distribution is then calculated by dividing the sum of each effect size multiplied by its weight by the sum of the weights (Hedges & Olkin, 1985; Shadish & Haddock, 1994). The standard error of the grand mean is simply the square root of the inverse of the sum of the weights, while the 95% confidence interval is calculated as the grand mean plus or minus 1.96 times the standard error. Random-effects results assume that differences in effect sizes result from both within- and between-grantee differences and that the intent of the analysis is to

generalize beyond the sample obtained. Because not all grantees contributed data to these analyses, random-effects models were adopted to estimate the impact of the Initiative on all grantees, and results are estimated using the DerSimonian and Laird (1986) method of moments.

To assess the probability that the grand mean calculated using the above procedures is well represented by the distribution of outcome scores obtained by grantees, meta-analysts use the  $Q$ -statistic (Hedges & Olkin, 1985).  $Q$  is an estimate of the variability of effect sizes in each distribution, and the  $p$  associated with the  $Q$  estimates the probability that the samples contributing to the estimate were drawn from a single underlying population.  $I^2$  further partitions the variance by estimating the proportion of variance that is associated with between-grantee differences (Higgins & Thompson 2002; Higgins, Thompson, Deeks, & Altman, 2003). In other words, “it is convenient to view  $I^2$  as a measure of *inconsistency* across the findings ... and not as a measure of the real variation across underlying true effects” (Borenstein, Hedges, Higgins, & Rothstein, 2009, p. 118, italics in the original).

*Meta-regression.* To test the influence of pre-grant environment, grant operations, and program characteristics in improving the seven outcomes, grantee outcomes were weighted by their method-of-moments-adjusted weight (Raudenbush, 2009) and regressed on five program characteristics after controlling for the influence of eight grant operations and four pre-grant environment control variables. Because we are generalizing to all grantees and not all grantees contributed data to each outcome analysis, random-effects regression modeling was used to estimate the influence of these control and independent variables on the amount of grantee change in outcomes (as represented by each grantee’s LOR). Random-effects models assume that there may not be a single underlying distribution represented by the contributing estimates but that measured and unmeasured variables may be influencing the results obtained by each site.

Therefore each measured outcome represents a unique distribution of possible outcomes. To account for this unknown variation, random-effects models estimate the random-effects variance component, recalculates the inverse variance weights with the random-effects component added, and refits the model (Lipsey & Wilson, 2001; Overton, 1998; Raudenbush & Bryk, 2001; Shadish & Haddock, 1994).

The lack of a comparison group and randomized assignment creates special challenges for what is essentially a natural experiment using observational data. It is therefore essential to control for pre-grant environmental factors (history of current partnership, poverty, grant recipient structure, and funding per targeted capita) before attributing differences in grantee performance to grant operations or grantee activities. Statistically, this was accomplished by using forced-entry regression to control for these influences prior to attributing differences in grantee performance to grant activities. In forced-entry regression, variables within logically defined blocks are entered in sequence, and variables that contribute significantly to the model at point of entry are retained in the model. Thus, the contribution of pre-grant environment variables is tested first. Next, models are refined by testing the additional contribution of grant operation variables. Lastly, the contribution of grant activities is tested by adding program characteristics variables to the models and assessing the improvement to each model (Cohen, J., & Cohen, P., 1983). The final models contain all variables that were significant when entered, and the contribution of each successive block of data can be tested by examining the additional variance explained once the block is entered. It can be noted that, while program theory determined when variable blocks were entered, the logic for retaining specific variables within each block was empirical; for each outcome, only those variables that notably improved the model were retained.

Because meta-analytic data are based on group-wise samples, degrees-of-freedom conventions in meta-regression are somewhat relaxed compared to those for conventional regression analysis. Nonetheless, given the limited degrees of freedom for testing models, variables within blocks were tested in order of the strength of their bivariate association with outcomes.

All analyses were conducted using Comprehensive Meta-analysis 2.2.055 and the meta-analysis and MetaReg macros for SPSS developed by David B. Wilson for regressing meta-analytic data. The macros used for these analyses are available on-line at:

<http://mason.gmu.edu/~dwilsonb/ma.html>.

### **3. Results**

Including outcomes comparing grantee performance over less than the 4 years possible dramatically increased the amount of data available for analysis (see Table 1). For example, 35 of the 59 grantees (59.3% of grantees) supplied both Year 3 and Baseline data for the outcome past 30-day alcohol use. Another 15 grantees provided past 30-day alcohol use data for Year 1 and Year 3 or for Year 2 and Year 3. Including these estimates brought the total number of grantees providing past 30-day alcohol use data to 50 of 59 grantees (84.7% of grantees), with 70.0 percent of the available data provided by Year 3 and Baseline estimates. Experienced violence had the highest and access to community-based mental health services outcomes had the lowest proportions of Year 3 and Baseline data, respectively. After including data from other waves of collection, access to community-based mental health services was the least commonly reported outcome, with estimates for 33 of 59 grantees (55.9% of grantees). Overall, 60.8 percent of the outcome data came from Year 3 and Baseline data; including other year comparisons increased the data available for analysis from 183 Year 3 and Baseline estimates to 301 outcome

estimates for the seven outcomes. Effect-size magnitude did not vary significantly or systematically by year of comparison for any of the outcomes.

<<<Insert Table 1 about here>>>

### *3.1 Outcomes by grantee*

To answer our first question, “do some grantees perform better across outcomes than other grantees?” we pooled outcomes measures evidence within each of the 57 grantees to estimate their average performance across outcomes and to determine whether the variation in outcomes was significant. Averaging within-grantee estimates using random-effects modeling (i.e., averaging each grantee’s results for the seven outcomes) shows that 48 of the 57 grantees providing data (84.2%) reported heterogeneous distributions of results (see Figure 1; also Appendix 1). In other words, their results for the different outcomes varied more than would be expected from sampling error (the average standard error across all grantees is 0.155; median = 0.132; Range = 0.029 – 0.418). Seven of the nine grantees reporting homogeneous distributions showed significant positive results, while the two remaining homogeneous distributions were submitted by grantees reporting significant negative distributions. Performance across outcomes varied more than would be expected from sampling error for every grantee showing no significant overall change in performance across outcomes over time.

<<<Insert Figure 1 about here>>>

Despite the diversity of performance on the seven outcomes, some patterns emerge in terms of the achievements of grantees in improving youth outcomes (see Table 2). Grantees that improved youth outcomes for one substance tended to be effective in reducing other youth substance use ( $r = .306 - .445, p < .05$ ). Past 30-day tobacco use was associated with improved access to community-based mental health services ( $r = .405, p < .05$ ), and reductions in past 30-

day marijuana use was associated with reduced perceived violence ( $r = .346, p < .05$ ). None of the other relations achieved significance, and several approach zero. Perhaps surprising, for the 37 grantees reporting both experienced violence and perceived violence, the correspondence between outcomes was quite low ( $r = .097, ns$ ). The relationship of access to school-based mental health services with access to community-based mental health services is larger but failed to reach significance ( $r = .253, ns$ ).

<<<Insert Table 2 about here>>>

Figure 2 provides another view of each grantee's results on the outcomes they reported (note that the graph has been trimmed to eliminate 3 extreme values). The lowest performing grantee, on average, occupies the top row of the graph, while the highest performing grantee, on average, occupies the bottom row. Perhaps the most striking observation is just how disparate within-grantee results are on the outcomes they reported. Of the 57 grantees who provided data, six (10.5%) show all positive results (one providing results that were all positive and significant), two (3.5%) reported all negative results, and no grantee reported all negative and significant findings. Also striking is the number of negative estimates reported by grantees. Fully 52.5 percent of the change scores, calculated from annually reported percentage estimates, are negative. Across all grantees and outcomes, 67 effect sizes (22.3%) are positive and significant, 28 are negative and significant (9.3%), and 206 (68.4%) show no significant change in outcome between periods of measurement.

<<<Insert Figure 2 about here>>>

### 3.2 *Main effects*

Overall, grantees were successful in meeting several of their outcomes. Based on the 95% confidence intervals, a somewhat more conservative test of significance than the standard

method (Payton, Greenstone, & Schenker, 2003; Schenker & Gentleman, 2001), three of the outcomes (access to school-based mental health services, access to community-based mental health services, and experienced violence) show significant and positive results under the assumptions of random-effects modeling. Converting the LORs to the d-scores more familiar to social scientists shows that, on average, over the periods of measurement, access to school-based mental health services increased significantly ( $d = 0.212$ , CI 0.106 – 0.318,  $I^2 = 0.0$ ) as did access to community-based mental health services ( $d = 0.148$ , CI 0.026 – 0.270,  $I^2 = 0.0$ ; see Figure 3). Significantly fewer students experienced violence ( $d = 0.057$ , CI 0.017 – 0.097,  $I^2 = 0.0$ ), and while no change in perceived risk of violence was observed ( $d = 0.010$ , CI -0.045 – 0.066,  $I^2 = 22.9$ ). Under the assumptions of random-effects meta-analysis, no change in past 30-day tobacco use ( $d = 0.031$ , CI -0.007 – 0.069,  $I^2 = 1.3$ ), past 30-day alcohol use ( $d = 0.004$ , CI -0.030 – 0.037,  $I^2 = 19.9$ ), or past 30-day marijuana use ( $d = -0.029$ , CI -0.082 – 0.024,  $I^2 = 0.0$ ) was observed. Converting these meta-analytic results to Rosenthal's BESD suggests that access to school-based mental health services increased 7.3 percent, access to community-based mental health services increased 4.1 percent, and 2.8 percent fewer students experienced violence (Rosenthal, 1994).

<<<Insert Figure 3 about here>>>

Significant heterogeneity is observed in three of the seven distributions (access to school- or community-based mental health services and experienced violence), but for each of these essentially all the variation is attributable to lack of precision in the original grantee estimates (large within-grantee variance). Past 30-day marijuana use likewise shows little between-study variability, but in this case the lack of between-study variance is attributable to  $Q$  being less than the degrees of freedom.

Across all outcomes, the SS/HS initiative was associated with a slight significant positive improvement in the seven student outcomes addressed by the initiative ( $d = 0.04$ , CI 0.001 – 0.088). There was, however, moderate between-outcome variation in the results obtained by grantees ( $I^2 = 44.9$ ).

### 3.3 *Random-effects meta-regression*

Under the assumptions of random-effects modeling, four of the seven outcomes showed insufficient between-grantee variability to justify modeling differences attributable to pre-grant, grant operations, or program characteristics. The proportion of variance attributable to between-grantee differences for the remaining three outcomes—past 30-day alcohol use, past 30-day tobacco use, and perceived risk of violence—could be described, at best, as low (Higgins et al., 2003). Nonetheless, between-grantee variation in these outcomes was modeled using random-effects meta-regression. As can be seen in Table 3, none of the hypothesized pre-grant, grant operations, or program characteristics variables predicted change in past 30-day alcohol use. The difference in change in past 30-day tobacco use is entirely predicted ( $R^2 = .997$ ) by a single pre-grant variable, with grantees that contained multiple local education agencies (LEA) showing greater improvement than grantees containing a single LEA. No pre-grant environment variables contributed significantly to the amount of change in perceived risk of violence. However, grantees with fewer non-required partners ( $\beta = -.307$ ) and grantees that placed higher value on partners' contributions ( $\beta = .331$ ) reported greater reduction in perceived risk of violence than did grantees who had more non-required partners who contributed less ( $R^2 = .993$ ).

<<<Insert Table 3 about here>>>

#### **4. Summary and Discussion**

The SS/HS Initiative is an unprecedented collaboration between the U.S. Departments of Health and Human Services, Education, and Justice. This evaluation assesses the impact of the SS/HS Initiative on seven student outcomes reported by the 57 grantees funded in 2005 and 2006.

Several limitations are evident in this evaluation. Most prominent is the lack of a control group against which to estimate impact. As a substitute, we adopted a comparative-effectiveness-research approach to examine how natural variation in grant operations and program characteristics outcomes were associated with grantee success in achieving outcomes. Also noteworthy is the decision to reduce grantee burden by allowing grantees to contribute data collected using extant data systems. Meta-analytic approaches were adopted to standardize those locally generated data and, with 52.5 percent of the effect sizes calculated from grantee-supplied data showing declines in youth outcomes, we are reassured that grantees did not attempt to bias the evaluation by selectively reporting outcomes that showed improvement. Limiting the utility of those data, however, was a lack of pre-grant and annual data for some measures due to variations in existing data collection infrastructures (e.g., reliance on bi-annual state data collection systems, time needed to establish data monitoring systems) and the Agency decision not to require performance data on all measures to be submitted during each reporting cycle. This reduced the data available for analysis and necessitated both imputing outcome data using best-estimates data when Year 3–Baseline data were not available and also required using random-effects modeling to generalize findings to grantees not submitting data.

Requiring grantees to provide performance estimates for all outcomes could improve future evaluations adopting this approach. With Census data the analyst has the option to adopt

fixed-effect models, which, while limiting inference to only those exposed to the Initiative, would have demonstrated that for those providing estimates the Initiative significantly improved five out of the seven outcomes. Fixed-effect models would have also allowed greater modeling of the variation in effects. Under the assumptions of fixed-effect modeling, between 50.9 and 97.1 percent of the heterogeneity (versus 0.0 – 22.9% under the assumptions of random-effects modeling) in outcome distributions is attributable to between-grantee variation.

Results from these grantees suggest that, while the SS/HS Initiative is associated with improvements in the lives of youths, the results for any given grantee vary broadly. With few exceptions, grantees reported disparate results across the outcomes they chose to report. Of the 57 grantees providing data, only 9 (15.8%) reported change scores across outcomes that might be considered representative of a single overall effect. Of those nine, only one showed significant and positive effects for all outcomes. For the remainder, data they provided suggest notably different levels of performance, with the majority showing improvements in some outcomes, while performance on other outcomes worsened or remained unchanged.

Given such variability in results, it is perhaps difficult to imagine that, using the data submitted to the national evaluation, a local evaluator would conclude that the initiative was improving youth outcomes. If true, this observation highlights the importance of large-scale evaluations for establishing effectiveness in real-world settings as the positive effect of the Initiative is only discernible at the meta-analytic level.

The diversity of within-grantee results may be attributable to the lack of a common liability for these groups of outcomes (cf. Jessor & Jessor, 1977). Although priority is given to grant applications that have comprehensive plans for addressing each of the six elements specified in the Initiative (Safe School Environments and Violence Prevention; Alcohol,

Tobacco, and Other Drug Prevention; Student Behavioral, Social, and Emotional Supports; Mental Health Services; Safe School Policies; and Early Childhood Social and Emotional Learning), grantees are nevertheless encouraged to focus their efforts on locally defined needs, gaps, or weaknesses. In other words, although comprehensive plans are encouraged, they are not a grant requirement.

Within the elements for which grantees submitted data, we found limited support for a common-liability argument. Grantees that were effective for one substance tended to reduce other substance use. Similarly, although change in access to school- and community-based mental health services is only modestly correlated, both are strongly positive. Differences in these outcomes may be attributable to choices in resource allocation, with some grantees leveraging the initiative to increase school- and others community-based mental health capacity.

Results from the two violence outcomes present a greater challenge to the assumption of common liability. For the 37 grantees that provided evidence for both measures, the correspondence between change in perceived violence and change in actually experiencing violence was almost zero, suggesting that reducing actual violence may not be sufficient for reducing perceived violence. It is not difficult to imagine that programs which increase a student's awareness of and attention to violence may increase a student's perceived risk of violence regardless of the actual rate of violence. Alternatively, it has also been suggested that perceived risk may increase the potential seriousness of violent episodes when they do occur (Webster, 1993). A singular severe violent episode may increase perceptions of risk in a way that multiple minor episodes do not.

Several factors militated against effectively modeling the correlates of grantee effectiveness. Methodologically, the desire to generalize from limited data and the disparate

sources of those data necessitated the use of random-effects modeling. After accounting for within-grantee variance, there was limited between-grantee variance available for modeling the possible reasons for grantee performance differences. More subversive for modeling common factors influencing success, however, is the substantive performance of grantees across outcomes. With few exceptions, the performance of grantees relative to one another varied greatly by outcome. Performance on outcomes differed significantly for most grantees, and their relative ranking in achieving those outcomes varied wildly. Although the limited between-group variance prevented systematically testing the consistency of grantee moderators across outcomes, the varied performance and disparate ranking of that performance suggests that different factors would be associated with success in achieving each of these outcomes. In other words, a common theory may not be sufficient for explaining how grantees improve different youth outcomes.

## **5. Conclusions**

Observing results at the grantee level, only seven grantees reported positive and significant grand mean results across the outcomes they measured, and only six grantees in this analysis reported all positive findings. Using these data, it is unlikely that many grantee local evaluators would conclude the Initiative was successful. Although the desire to generalize findings to all grantees limited the variance available for modeling and created a higher standard for documenting effectiveness, taken as a whole, the SS/HS Initiative is associated with improved outcomes for youth. Modest to significant improvement is observed in four of seven outcomes, and the overall impact of the Initiative is positive and significant. This evaluation demonstrates the value of rigorous cross-site analysis for estimating the effectiveness of large-scale Initiatives and how locally collected data can be standardized and included in such

analyses.

### **Acknowledgements**

Funding for this study was provided by the Center for Mental Health Services, Substance Abuse and Mental Health Services Administration within the U.S. Department of Health and Human Services. The content of this article represents the views of the authors and does not reflect the views of the Substance Abuse and Mental Health Services Administration or the U.S. Department of Health or Human Services.

## References

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley & Sons, Ltd.
- Butterfoss, F. D. (2007). *Coalitions and partnerships in community health*. San Francisco: Jossey-Bass.
- Butterfoss, F. D., Goodman, R. M., & Wandersman, A. (1993). Community coalitions for health promotion and disease prevention. *Health Education Research*, 8(3), 315-330.
- Chen, J., & Shao, J. (2000). Nearest neighbor imputation for survey data. *Journal of Official Statistics*, 16, 113-131.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral Sciences* (2<sup>nd</sup> Edition). Hillsdale, NJ: Lawrence Erlbaum Associates.
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177-188.
- Epstein, J. L. (1995). School/family/community partnerships: Caring for the children we share. *Phi Delta Kappan*, 76, 701-712.
- Flannery, D. J., Singer, M. I., Van Dulman, M., Kretschmar, J., & Belliston, L. (2007). Exposure to violence, mental health and violent behavior. In D. J. Flannery, A. T. Vazsonyi, & I. Waldman (Eds.), *The Cambridge handbook on violent behavior*. Cambridge: Cambridge University Press. (pp. 306-321).
- Fleiss, J. L. & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (2nd ed), New York: Russell Sage. (pp. 237-253).

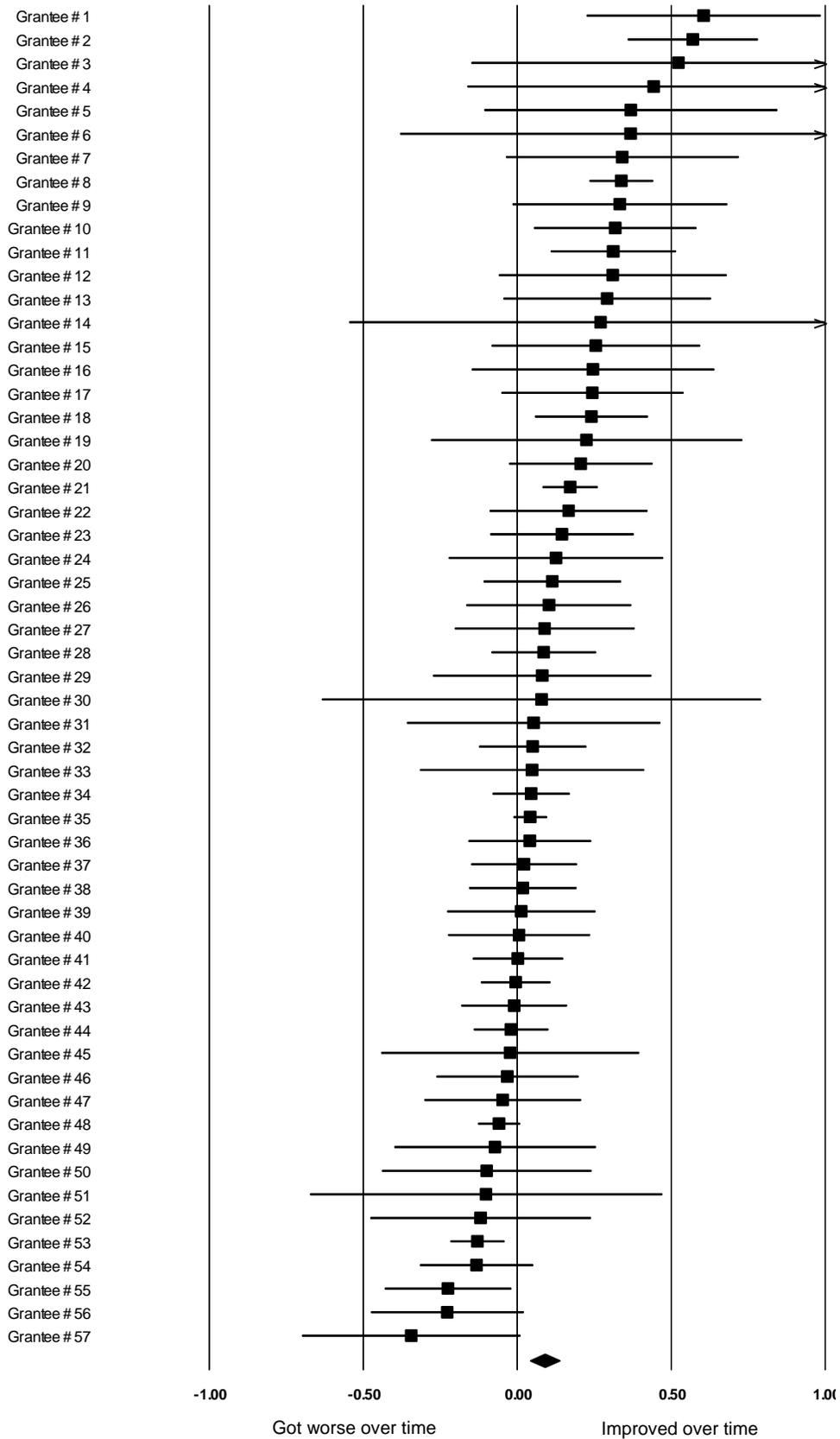
- Hasselblad, V. & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin*, 117(1), 167-178.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.
- Higgins J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539-1558.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557-560.
- Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. New York: Academic Press.
- Kodjo, C. M., Auinger, P., & Ryan, S. A. (2003). Demographic, intrinsic, and extrinsic factors associated with weapon carrying at school. *Archives of Pediatrics and Adolescent Medicine*, 157(1):96-103.
- Komro, K. A., Williams, C. L., Forster, J. L., Perry, C. L., Farbakhsh, K. & Stigler, M. H. (2000). The relationship between adolescent alcohol use and delinquent and violent behaviors. *Journal of Child and Adolescent Substance Abuse*, 9(2), 13-28.
- Kreuter, M. W., Lezin, N. A., & Young, L. A. (2000). Evaluating community-based collaborative mechanisms: Implications for practitioners. *Health Promotion Practice*, 1(1), 49-63.
- Lannie, A. L., & McCurdy, B. L. (2007). Preventing disruptive behavior in the urban classroom: Effects of the Good Behavior Game on student and teacher behavior. *Education and Treatment of Children*, 30(1), 85-98.

- Lasker, R. D., & Weiss, E. S. (2003a). Broadening participation in community problem solving: A multidisciplinary model to support collaborative practice and research. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 80, 14-59.
- Lasker, R. D., & Weiss, E. S. (2003b). Creating partnership synergy: The critical role of community stakeholders. *Journal of Health and Human Services Administration*, 26, 119-139.
- Lasker, R. D., Weiss, E. S., & Miller, R. (2001). Partnership synergy: A practical framework for studying and strengthening the collaborative advantage. *Milbank Quarterly*, 79(2), 179-205.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Applied Social Research Methods Series, Vol. 49. Thousand Oaks, CA: Sage.
- Mitchell, S. M., & Shortell, S. M. (2000). The governance and management of effective community health partnerships: A typology for research, policy, and practice. *The Milbank Quarterly*, 78(2), 241-289.
- Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3, 354-379.
- Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science*, 3, 34. Available online: [www.insectscience.org/3.34](http://www.insectscience.org/3.34). Accessed 11/23/2009.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon & Schuster.

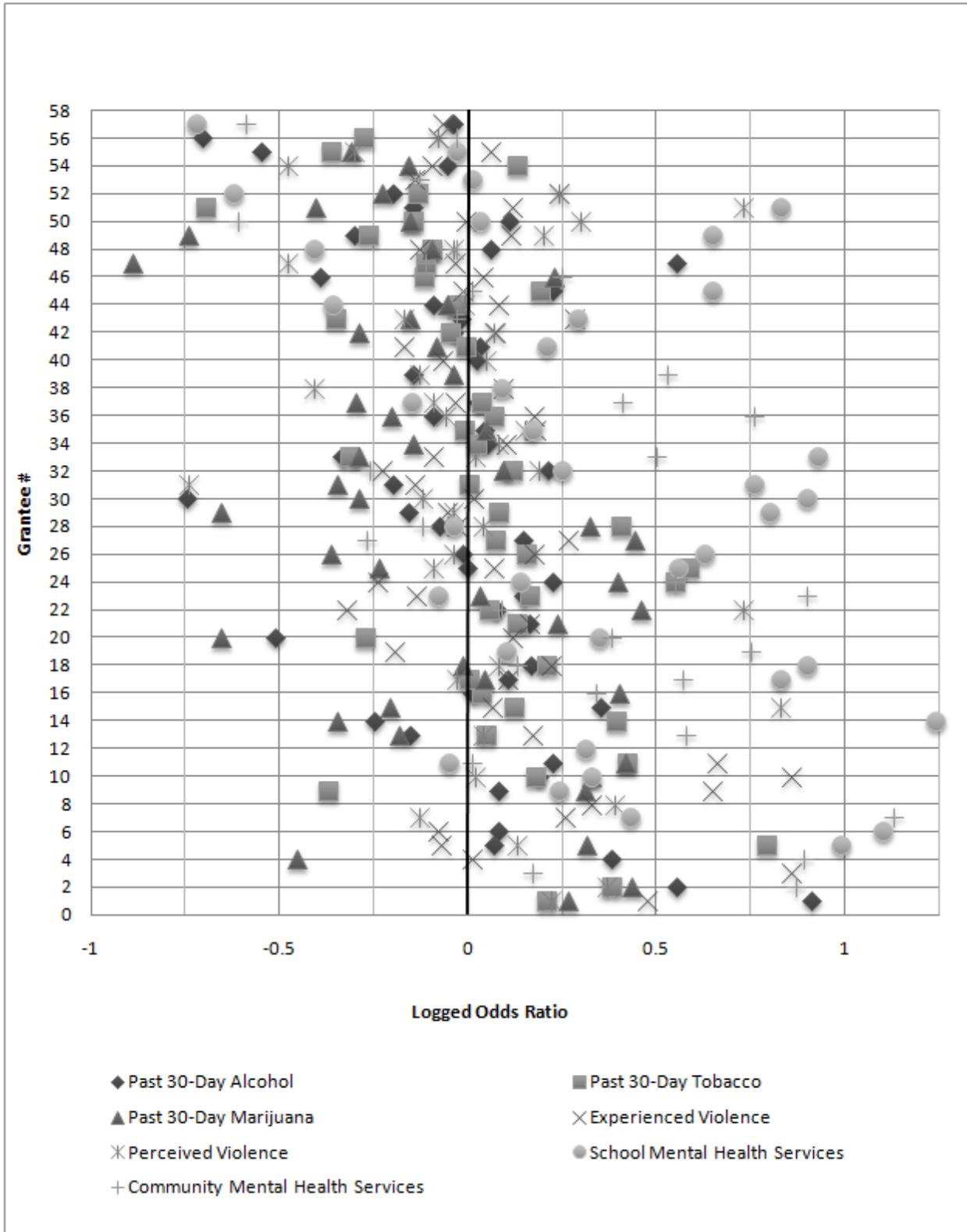
- Rancourt, E., Sarndal, C. E., & Lee, H. (1994). Estimation of the variance in the presence of nearest neighbor imputation. *Proceedings of the Section on Survey Research Methods. American Statistical Association*, 888-893.
- Raudenbush, S. W. (2009). Analyzing effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (2nd ed), (pp. 295-315). New York: Russell Sage.
- Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Robers, S., Zhang, J., & Truman, J. (2010). *Indicators of school crime and safety: 2010*. NCES 2011-002/NCJ 230812. Washington, DC: National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.
- Rollison, J., Hill, G., Yu, P., & Murray, S. (2011). Evaluation of a complex, multisite, multilevel grants initiative. *Evaluation and Program Planning*. (forthcoming, this issue)
- Rosenthal, R. (1994). Statistically describing and combining studies. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage. (pp. 231-244)
- Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, 55, 182-186.

- Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In: H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation; pp. 261-281.
- Tolan, P. H., & Gorman-Smith, D. (2002). What violence prevention can tell us about developmental psychopathology. *Development and Psychopathology, 14*, 713-729.
- Wandersman, A., Goodman, R., & Butterfoss, F. (1997). Understanding coalitions and how they operate. In M. Minkler (Ed.), *Community organizing and community building for health*. New Brunswick, NJ: Rutgers University Press.
- Webster, D. W. (1993). The unconvincing case for school-based conflict resolution programs for adolescents. *Health Affairs, 12*(4), 126-141. doi: 10.1377/hlthaff.12.4.126.
- Weissberg, R. P., Kumpfer, K. L., & Seligman, M. E. P. (2003). Prevention that works for children and youth. *American Psychologist, 58*, 425-432.

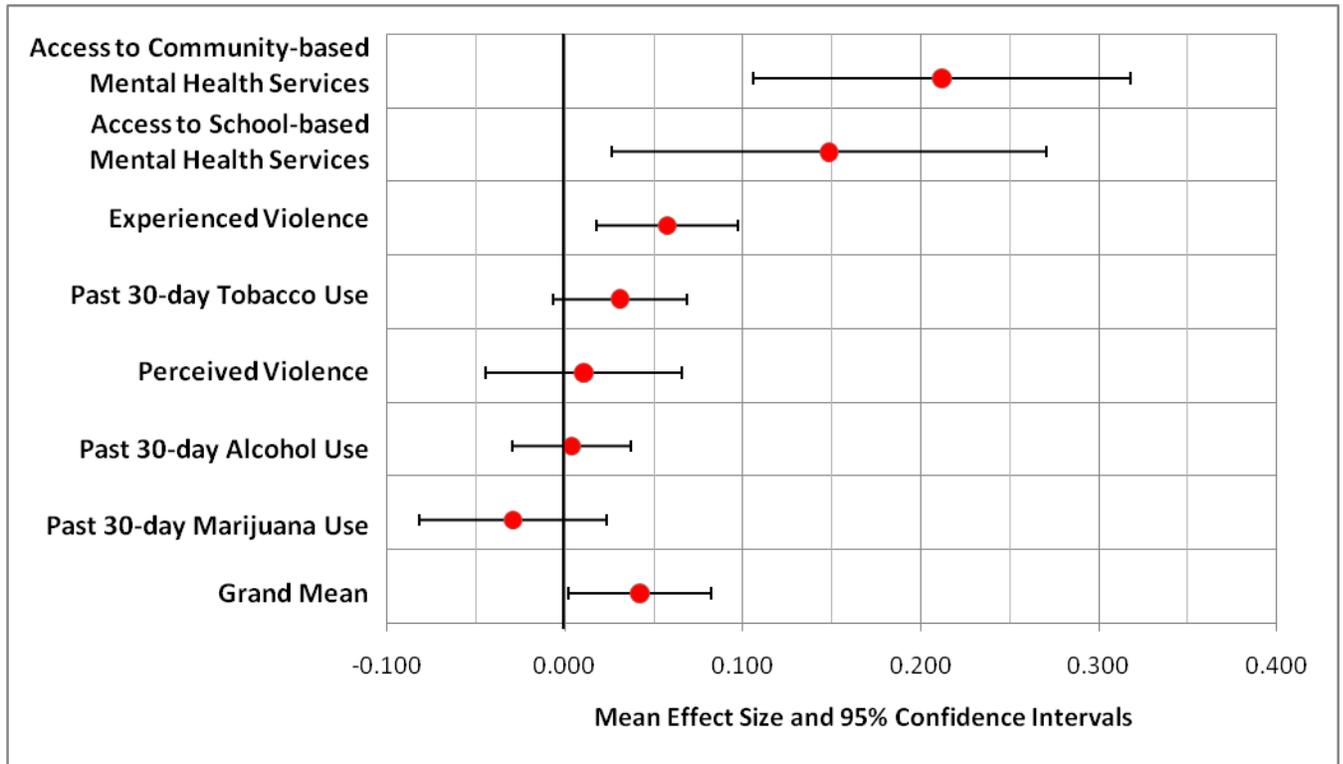
**Figure 1: Random-effect mean logged odds ratios across all reported outcomes: By grantee**



**Figure 2: Logged Odds Ratios by Grantee by Outcome**



**Figure 3: Random-Effects Mean d-scores for Seven Outcomes**



**Table 1: Source and Frequency of Nearest-Neighbor Estimates**

<b>Outcome</b>	<b>Year 3 and Baseline</b>	<b>Percent of 59 Grantees with Year 3 and Baseline Data</b>	<b>Year 3 and Year 1</b>	<b>Year 3 and Year 2</b>	<b>Total Number of Estimates</b>	<b>Percent of Data from Year 3 and Baseline</b>
Past 30-day Alcohol Use	35	59.3%	11	4	50	70.0%
Past 30-day Tobacco Use	33	55.9%	7	4	44	75.0%
Past 30-day Marijuana Use	32	54.2%	10	3	45	71.1%
Experienced Violence	36	61.0%	14	2	52	69.2%
Perceived Violence	21	35.6%	14	4	39	53.8%
Access to School-based Mental Health Services	14	23.7%	15	9	38	36.8%
Access to Community-based Mental Health Services	12	20.3%	10	11	33	36.4%
Total Number of Estimates	183		81	37	301	60.8%

**Table 2: Correlations between Outcomes**

	Past 30 Day Tobacco (N)	Past 30 Day Marijuana (N)	Experienced Violence (N)	Perceived Risk of Violence (N)	Access to School-based Mental Health Services (N)	Access to Community-based Mental Health Services (N)
Past 30 Day Alcohol	<b>.435</b> (44)	<b>.360</b> (45)	.201 (46)	.157 (36)	-.181 (33)	.219 (29)
Past 30 Day Tobacco		<b>.445</b> (42)	-.056 (41)	-.080 (33)	.119 (30)	<b>.405</b> (26)
Past 30 Day Marijuana			.240 (41)	<b>.346</b> (33)	-.270 (30)	.044 (26)
Experienced Violence				.096 (37)	.014 (36)	.055 (30)
Perceived Risk of Violence					-.039 (26)	-.315 (21)
Access to School-based Mental Health Services						.253 (21)

**Bold** = Significant at .05

**Table 3: Final Random Effects Meta-Regression Models**

	Past 30-Day Alcohol (I <sup>2</sup> = 19.9)	Past 30-Day Tobacco (I <sup>2</sup> = 1.3)	Perceived Risk of Violence (I <sup>2</sup> = 22.9)
<b>Number of observations</b>	50	44	39
<b>Pre-grant Environment Variables</b>			
Having multiple local educational agencies		<b>0.245</b>	
<b>Percent of Available Variance Explained, Control Variables only</b>	0.0%	<b>99.7%</b>	0.0%
<b>Grant Operations Variables</b>			
Level and type of partnership contributions - Year 2			<b>0.331</b>
Number of non-required partners - Year 2			<b>-0.307</b>
<b>Percent of Available Variance Explained, Pre-grant and Grant Operations Variables</b>	0.0%	<b>99.7%</b>	<b>99.3%</b>

**Bold** = Significant at p < .05

## Appendix 1: Results by Grantee

	<b>Number of outcomes</b>	<b>Mean LOR</b>	<b>Lower 95% CI</b>	<b>Upper 95% CI</b>	<b>z</b>	<b>p</b>
Grantee # 1*	6	0.605	0.223	0.987	3.103	0.002
Grantee # 2*	5	0.570	0.355	0.784	5.209	0.000
Grantee # 3	2	0.523	-0.151	1.197	1.521	0.128
Grantee # 4	4	0.443	-0.166	1.051	1.426	0.154
Grantee # 5	6	0.369	-0.109	0.847	1.513	0.130
Grantee # 6	3	0.368	-0.383	1.118	0.960	0.337
Grantee # 7	4	0.341	-0.040	0.721	1.753	0.080
Grantee # 8*	2	0.338	0.231	0.445	6.206	0.000
Grantee # 9	5	0.333	-0.019	0.685	1.856	0.063
Grantee # 10*	6	0.318	0.051	0.585	2.334	0.020
Grantee # 11*	6	0.312	0.107	0.518	2.979	0.003
Grantee # 12	1	0.310	-0.062	0.682	1.632	0.103
Grantee # 13	7	0.292	-0.047	0.631	1.690	0.091
Grantee # 14	4	0.271	-0.549	1.090	0.648	0.517
Grantee # 15	5	0.255	-0.085	0.595	1.469	0.142
Grantee # 16	4	0.246	-0.150	0.641	1.219	0.223
Grantee # 17	7	0.244	-0.053	0.541	1.607	0.108
Grantee # 18*	7	0.241	0.054	0.428	2.528	0.011
Grantee # 19	3	0.225	-0.283	0.734	0.868	0.386
Grantee # 20	6	0.206	-0.028	0.441	1.722	0.085
Grantee # 21*	4	0.172	0.079	0.264	3.630	0.000
Grantee # 22	6	0.167	-0.091	0.425	1.268	0.205
Grantee # 23	6	0.145	-0.090	0.381	1.209	0.227
Grantee # 24	6	0.126	-0.225	0.477	0.701	0.483
Grantee # 25	6	0.114	-0.111	0.339	0.995	0.320
Grantee # 26	6	0.103	-0.167	0.374	0.749	0.454
Grantee # 27	5	0.089	-0.204	0.383	0.597	0.550
Grantee # 28	7	0.086	-0.087	0.259	0.977	0.328
Grantee # 29	6	0.081	-0.277	0.438	0.442	0.658
Grantee # 30	5	0.079	-0.636	0.794	0.216	0.829
Grantee # 31	7	0.053	-0.362	0.467	0.250	0.803
Grantee # 32	7	0.050	-0.128	0.227	0.549	0.583
Grantee # 33	7	0.048	-0.319	0.414	0.254	0.800
Grantee # 34	5	0.045	-0.082	0.171	0.696	0.487
Grantee # 35	6	0.042	-0.016	0.099	1.424	0.154
Grantee # 36	6	0.041	-0.160	0.242	0.399	0.690
Grantee # 37	7	0.022	-0.153	0.197	0.246	0.806
Grantee # 38	3	0.018	-0.158	0.194	0.203	0.839
Grantee # 39	4	0.013	-0.230	0.257	0.106	0.916
Grantee # 40	3	0.006	-0.226	0.239	0.054	0.957
Grantee # 41	5	0.002	-0.148	0.152	0.027	0.978

Grantee # 42	5	-0.005	-0.120	0.111	-0.081	0.935
Grantee # 43	7	-0.010	-0.183	0.164	-0.108	0.914
Grantee # 44	6	-0.020	-0.143	0.103	-0.321	0.748
Grantee # 45	6	-0.023	-0.444	0.399	-0.106	0.916
Grantee # 46	5	-0.032	-0.265	0.200	-0.272	0.786
Grantee # 47	6	-0.047	-0.303	0.209	-0.359	0.720
Grantee # 48	7	-0.059	-0.130	0.011	-1.641	0.101
Grantee # 49	6	-0.072	-0.401	0.258	-0.425	0.671
Grantee # 50	7	-0.099	-0.441	0.244	-0.563	0.573
Grantee # 51	7	-0.101	-0.675	0.474	-0.343	0.732
Grantee # 52	6	-0.119	-0.479	0.241	-0.649	0.517
Grantee # 53*	3	-0.129	-0.218	-0.039	-2.812	0.005
Grantee # 54	5	-0.132	-0.318	0.054	-1.391	0.164
Grantee # 55*	6	-0.225	-0.433	-0.016	-2.115	0.034
Grantee # 56	5	-0.227	-0.477	0.023	-1.778	0.075
Grantee # 57	4	-0.344	-0.700	0.013	-1.890	0.059

Key: Mean LOR = Random effects averaged logged odds ratio, Lower 95% CI = Lower 95% confidence interval, Upper 95% CI = Upper 95% confidence interval, z = z-score, p. = probability that the distribution is heterogeneous. \* = Grantee results do not vary more than would be expected from chance

## **Author Vitae**

**James H. Derzon**, Ph.D., of Battelle Centers for Public Health Research and Evaluation, is a nationally recognized expert in prevention science and the impact of research methods and procedures on study findings. For over 19 years Dr. Derzon has used meta-analytic techniques and methods to document the evidence surrounding risk and protective factors for substance use, antisocial behavior, and violence and the effectiveness of prevention programs and policies in reducing those behaviors. He has conducted studies on the quality and use of outcome and performance measures for documenting program and agency impact and on how service delivery options impact intervention effectiveness. Much of Dr. Derzon's work has been presented in the context of describing how service provision options and methodological choices of evaluators influence study findings.

**Ping Yu**, Ph.D., of Battelle Centers for Public Health Research and Evaluation, has more than 25 years of experience in community- and school-based substance abuse research, domestic and international program evaluation, and management of large-scale, multi-year, and multi-million-dollar projects. He currently serves as the Project Director for Battelle on the national evaluation of the Safe Schools/Healthy Students Initiative. Dr. Yu has managed a portfolio of research and evaluation projects worth more than \$60 million across a wide range of government and private sector clients. He has designed and directed a number of large-scale studies that use both qualitative and quantitative methods to assess the effectiveness of community- and school-based prevention or treatment initiatives. Dr. Yu's contributions have received broad peer recognition in the form of published articles and participation in such forums as the National Committee on Asian and Pacific Islander Americans Health Research convened by the Director of Minority Health at the National Institutes of Health.

**Bruce Ellis**, M.S., of Battelle Centers for Public Health Research and Evaluation, has more than 30 years of experience in biomedical and health research, statistical analysis, and substance abuse research. He has served as a Senior Analyst on the national cross-site evaluation of the Safe Schools/Healthy Students Initiative for three years, focusing on the development of hierarchical linear models of the Initiative's survey data and on meta-analyses of Government Performance and Results Act outcomes data. Previously, he has played key roles in the analysis of large cross-site evaluations, such as the Residential Women with Children/Pregnant and Post-Partum Women (RWC-PPW) program of the Center for Substance Abuse Treatment, and the Substance Abuse and Mental Health Services Administration's National Treatment Improvement Evaluation Study.

**Sharon Xiong**, M.S., of Battelle Centers for Public Health Research and Evaluation, is an experienced researcher who has worked on the national cross-site evaluation of the Safe Schools/Healthy Students Initiative for four years. She has played an extensive role in the collection and quality control of Government Performance and Results Act outcomes data, in the management and editing of school- and project-level survey data, and in support of numerous analyses such as this research. Prior to the evaluation of the Safe Schools/Healthy Students Initiative, she has supported other large, multi-site program evaluations such as the Substance Abuse and Mental Health Services Administration's Center for Substance Abuse Prevention's State Incentive Grants program.

**Carmen Arroyo, Ph.D.**, currently Special Expert in Evaluation for the Division of Prevention, Traumatic Stress, and Special Programs at the Substance Abuse and Mental Health Services Administration, has over 20 years of experience in research and evaluation of health and education programs. Previously she worked as a Principal Research Scientist at the American Institutes for Research, where she helped federal agencies, including the Department of Education, the Office of Juvenile Justice, the Centers for Disease Control and Prevention, and the Agency for International Development to set standards for the evaluation of evidence-based education and health interventions. She has been Principle Investigator on projects related to educational access and attainment, comprehensive community health initiatives, maternal and child health, adolescent health, school-based health programs, HIV/AIDS services for poor and ethnically diverse communities, and the integration of mental health services within primary care. Dr. Arroyo holds a Ph.D. in Social and Developmental Psychology from Yale University.

**Gary Hill, Ph.D.**, of MANILA Consulting Group, Inc., has nearly 40 years of increasing responsibility in the fields of health policy, program evaluation, project management, and the design and development of automated information systems. He currently serves as the Project Director for the Safe Schools/Healthy Students national cross-site evaluation. Previously, Dr. Hill was Division Director at a large federal government contracting firm, overseeing contracts providing, for example, program evaluation, cost-benefit analyses, and reports to Congress. Dr. Hill was Project Director for the evaluation of the Robert Wood Johnson Foundation's After School Program. He was also Project Director for the Department of State's International Demand Reduction Program, a \$15 million, multiyear effort designed to coordinate implementation and evaluation of prevention programs across the world.

**Julia Rollison, M.P.H.**, of MANILA Consulting Group, Inc., currently serves as a Senior Analyst for the national cross-site evaluation of the Safe Schools/Healthy Students Initiative. She manages quantitative and qualitative data collection and coordinates data analysis efforts on key outcome measures, specifically those associated with Government Performance and Results Act requirements. She is responsible for ensuring the quality and integrity of data collected and creating and maintaining cross-site data repositories. Ms. Rollison has expertise in program evaluation and performance measurement from past evaluations, such as the 5-year, multicity study on the outcomes of financing and organizing after-school activities by the Robert Wood Johnson Foundation, and the cross-site evaluations of 35 initiatives for the Substance Abuse and Mental Health Services Administration's Fetal Alcohol Spectrum Disorders Center for Excellence.