The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title:	**Assessing Different Levels and Dosages of the Shifting Boundaries Intervention to Prevent Youth Dating Violence in New York City Middle Schools: A Randomized Control Trial**
Author(s):	**Bruce G. Taylor, Ph.D., Elizabeth A. Mumford, Ph.D., Weiwei Liu, Ph.D., Nan D. Stein, Ed.D.**
Document No.:	**249587**
Date Received:	**January 2016**
Award Number:	**2010-MU-MU-0008**

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this federally funded grant report available electronically.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Final Report

Assessing different levels and dosages of the Shifting Boundaries intervention to prevent youth dating violence in New York City middle schools: A randomized control trial*

December 18, 2015
Report submitted to the National Institute of Justice
Grant # 2010-MU-MU-0008

Bruce G. Taylor, Ph.D., NORC at the University of Chicago**
Elizabeth A. Mumford, Ph.D., NORC at the University of Chicago
Weiwei Liu, Ph.D., NORC at the University of Chicago
Nan D. Stein, Ed.D., Wellesley College, Center for Research on Women, Wellesley Centers for Women

* This study ("A dating violence prevention program for each grade in middle school: A longitudinal multi-level experiment") was funded by the National Institute of Justice (Grant # 2010-MU-MU-0008). The views expressed are those of the authors and do not necessarily represent the views or the official position of the National Institute of Justice or any other organization.

**** Address correspondence to:**
Bruce G. Taylor, Senior Fellow, NORC at the University of Chicago, 4350 East West Highway, Room 709, Bethesda, MD 20814, taylor-bruce@norc.org

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Acknowledgements

This study was funded by the National Institute of Justice (Grant # 2010-MU-MU-0008) and we are grateful for this funding and how it allowed us to continue our testing of the *Shifting Boundaries* program. We thank our Project Officer Dr. Carrie Mulford and Senior Grants Management Specialist Laurie Bright for their encouragement and support. We would like to express our gratitude to the principals, staff, and students at all of our participating schools, for this research would not have been possible without their help. Thanks also to those at the New York City Department of Education who helped to make this research possible. In particular, Scott Bloom, LCSW (Director of School Mental Health Services, Office of School Health, New York City Department of Education) was instrumental with helping our team locate schools interested in participating in this research. We would like to acknowledge the rest of the Shifting Boundaries team, including Keith Towery for all his work on managing the project logistics.

Conflict of Interest: The authors have no conflicts of interest to disclose.

Financial Disclosure: The authors have no financial relationships relevant to this article to disclose.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table of Contents

ACKNOWLEDGEMENTS **2**

ABSTRACT **5**

EXECUTIVE SUMMARY **7**

INTRODUCTION **14**

LITERATURE REVIEW **16**

DATING VIOLENCE AND SEXUAL HARASSMENT: SCOPE OF THE PROBLEM 16
THE NATURE OF DV/H 17
Onset and developmental pathways. *17*
Consequences associated with DV/H. *17*
SCHOOL-BASED VIOLENCE PREVENTION PROGRAMS 19
PRIOR SHIFTING BOUNDARIES (SB) EXPERIMENTS 21
THE CURRENT STUDY 24

METHODS **25**

DESCRIPTION OF INTERVENTIONS 25
Classroom-based intervention (SBC) *26*
School (building-level) based intervention (SBS) *28*
RESEARCH SITE LOCATION 29
EXPERIMENTAL DESIGN 29
CHALLENGES IN IMPLEMENTING OUR EXPERIMENT: 30
STUDENT SURVEY ADMINISTRATION 34
MEASURES 35
Sexual and physical violence victimization and perpetration *35*
THEORETICAL FRAMEWORK FOR INTERVENTIONS AND STUDY HYPOTHESES 36
DATA ANALYTIC ISSUES 38
Statistical power *38*
Pre-treatment study arm comparison *39*
Attrition analyses: *41*

RESULTS **43**

DESCRIPTIVE STATISTICS ON THE SAMPLE 43
Demographics *43*
Prior education in violence prevention *44*
Prior relationship experience *44*
Prior experience of victimization *44*

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Prior experience of perpetration *44*
OUTCOME MODELS 45
Six-month follow-up outcome data *46*
Twelve-month follow-up outcome data *49*

DISCUSSION **52**

FEW DIFFERENCES ACROSS THE TREATMENT GROUPS. 53
RESISTANCE FROM SCHOOLS AND ITS EFFECTS ON STUDY. 55
LIMITATIONS 56
IMPLICATIONS 58

CONCLUSION **61**

REFERENCES **62**

APPENDICES **76**

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Abstract

In this report we present the results of an experimental evaluation based on a randomized controlled trial (RCT) of a dating violence and sexual harassment (DV/H) prevention program — called the *Shifting Boundaries* (SB) Program. We randomly assigned 23 public middle schools in New York City to one of four treatment conditions of SB varying by dosage and saturation levels. The project includes a baseline and two follow-up surveys with 6th, 7th and 8th grade students to assess short to medium term impact on rates of DV/H. The intervention we tested had two main components. First, we had an SB classroom curriculum (SBC), covering the consequences for perpetrators of DV/H, laws and penalties for DV/H, and respectful relationships. Second, we had an SB school (building-level) (SBS) component which included the use of school-based restraining orders, higher levels of faculty and security presence in areas identified through student mapping of safe/unsafe "hot spots," and the use of posters to increase awareness and reporting of DV/H. We examined (1) the effects of saturating a school environment by providing the SB intervention to all three middle school grades compared to only two grades or one grade and (2) the effects of two dosages of SB across two years compared to one dosage of the SB intervention across one year.

Participating students ranged in age from 10 to 15, with a nearly 50% split between boys and girls. Our sample was 26% Hispanic, 37% African American, 16% Asian, 14% White and 7% "other." Over 40% of the sample had prior experience with a violence prevention educational program. Nearly half reported at least one experience of being in a dating relationship. At baseline, about one in five respondents reported having ever been the victim of any physical dating violence, with a similar number reporting perpetrating any physical dating violence. One in ten respondents reported having been the victim of any sexual dating violence ever (6.4% for perpetration of this act). Almost 60% of the sample reported having ever been the victim of any physical peer violence at some point in time (45% perpetration), and 18.1% were ever the victim of sexual peer violence (8% perpetration). Also, 49% reported experiencing sexual harassment (SH) at some point in time (23% perpetration).

Our overall results indicate that providing the SB treatment to only one grade level in middle school does just as well in terms of peer violence and dating violence outcomes as a more saturated process of treating multiple grades. At both the 6-month and the 12-month assessments, however, there was evidence that additional saturation beyond one grade is associated with reductions in sexual harassment victimization. Schools that delivered SB to both 6th and 7th graders (compared to just 6th graders) showed reductions SH victimization reports at 6 months post treatment, an effect that was still significant at the 12-month assessment. Also at 12 months post treatment, schools that delivered SB to all middle school grades (6th – 8th) showed reductions in self-report of SH victimization.

However, we also found that greater saturation of the SB program (delivered to 6th & 7th graders or to all three grades levels) was unexpectedly associated with more reported perpetration of sexual violence against peers at 12 month post treatment compared to the 6th grade only group, a finding in contrast with the additional borderline statistically significant findings ($p<.10$) at the 6-month assessment suggesting that receiving SB saturation for two grades rather than only one was associated with reduced frequency of peer physical victimization frequency and peer sexual violence perpetration.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

There were no results indicating that offering the SB program to a grade of students in two successive years (the 6th grade longitudinal design) resulted in statistically differential effects ($p < .05$) compared to a one-time dosage of SB in 6th grade. However, one borderline ($p < .10$) statistically significant effect (SB program delivered to 6th graders in year 1 and again to the same students, as 7th graders, in year 2 was associated with less SH victimization frequency compared to the 6th grade only intervention) highlights the potential potency of multiple dosages of the SB program for SH prevention work.

These results largely support a minimalistic approach, in that SB effectiveness for peer and DV/H outcomes may be achieved by delivery to only one grade level in middle schools. However, taking these results in the context of our earlier work (NYC-1), there is a rationale for considering saturated delivery of the school wide (SBS) component of SB. In earlier research, SBS was effective at reducing DV/H outcome independent of the classroom curriculum (SBC). Because the SBS program can be introduced to an entire middle school at low-cost, and our current research shows positive effects of exposing more than just a single grade to the SB program, these results taken together suggest policy and administrative consideration of a saturated delivery of the SBS program.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Executive Summary

Teen dating violence and sexual harassment (DV/H) (including "interpersonal" or "gender" violence or adolescent relationship abuse)[1] have been recognized as serious and persistent problems in the public health and violence prevention fields (Centers for Disease Control and Prevention, 2012a, 2014; Jouriles, Garrido, Rosenfield, & McDonald, 2009; Mulford & Giordano, 2008a; National Institute of Justice, 2011; Shanklin, Brener, McManus, Kinchen, & Kann, 2007). Studies reveal that over half of all teens are victims of teen dating violence (TDV) (Foshee, 1996; Hickman, Jaycox, & Aranoff, 2004; Jouriles, Platt, & McDonald, 2009; Malik, Sorenson, & Aneshensel, 1997; O'Keefe, 1997; Taylor & Mumford, 2014 e-pub ahead of print) and most have experienced sexual harassment (Hill & Kearl, 2011). DV/H can lead to serious injuries for victims, poorer mental/physical health, more "high-risk"/deviant behavior, and increased school avoidance (Gruber & Fineran, 2008; Howard, Wang, & Yan, 2007a, 2007b).

Background on DV/H Interventions

Despite the grim news on the vast prevalence and far reaching impact of DV/H, over the past decade a number of interventions have been developed to prevent DV/H and/or lower its reoccurrence. Also, rigorous research has been conducted on the effectiveness of DV/H prevention programs (Foshee & Reyes, 2009; Jaycox et al., 2006; Taylor, 2010; Wolfe et al., 2009), and a number of these studies have shown positive results. However, these studies are few and generally address only 8th and/or 9th grade or older students.

The study we report on in this final report builds on two earlier experiments of the *Shifting Boundaries* (SB) intervention. In 2005, some of the authors of this report (Taylor & Stein, Ohio 2005-2007) conducted one of the first experimental evaluations of a primary prevention program addressing DV/H for sixth and seventh grade students in suburban middle schools bordering Cleveland. The research team randomly assigned 123 study classrooms to one of three conditions: (1) An interaction-based curriculum, (2) a law and justice curriculum or a control group. This earlier research confirmed that DV/H reductions could be achieved with middle school prevention programming (Taylor, Stein, Mack, Horwood, & Burden, 2008; Taylor, 2010). From 2008 to 2010, in New York City (NYC) middle schools, the team conducted a second experiment (Taylor, Stein, Mumford, & Woods, 2013a). In this study (hereafter referred to as "NYC-1"), the researchers added a building-wide intervention component (Shifting Boundaries Schoolwide; SBS) to go along with the most effective components of the Cleveland classroom-based interventions to form a four-celled experiment testing the effectiveness of the

[1] In this report, we use the term *dating violence and harassment* (DV/H) to represent physical, emotional, or sexual abuse within a dating relationship, the definition that CDC uses for teen dating violence (TDV) (Centers for Disease Control and Prevention, Accessed 2/11/11). More broadly, this problem has also been referred to as gendered adolescent interpersonal aggression (GAIA) (Smith, White, & Moracco, 2009) Where cited studies used the term TDV, we also follow the language of the original research. We also will use the term DV/H to cover youth sexual harassment (SH). We follow the American Association of University Women's (AAUW) definition of SH. SH is unwelcome conduct of a sexual nature in person or through electronic means, which can include unwelcome sexual advances, requests for sexual favors, or other verbal, nonverbal, or physical conduct of a sexual nature (Hill & Kearl, 2011).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

classroom SB curriculum (SBC), SBS, both (SBC+SBS), or neither. Our team randomly assigned 30 public middle schools to one of four conditions: (1) receive the SBS and SBC interventions, (2) receive SBS only, (3) receive SBC only or (4) control group (in which schools/students experienced their normal class schedule, without receiving any elements of the classroom or building-level interventions). The main findings from this study were that the building-only and the both (classroom lessons with building-wide) interventions were effective at reducing dating and peer violence victimization and perpetration (Taylor et al., 2013a).

Overall, the SBS intervention and the SBC+SBS intervention were effective at reducing DV/H. The success of the "building only" intervention is particularly important because it can be implemented with very few extra costs to schools. While some important findings were produced in the NYC-1 results, a couple of key questions remained. First, whether the SB program was of sufficient dosage to produce sustained effects post intervention beyond the six months demonstrated in NYC-1. Second, if schools could implement the SB program in just one grade to conserve resources but still achieve DV/H reduction effects. While the Safe Dates evaluation (Foshee, Bauman, et al., 2004) had assessed the question of dosage (and found no additional TDV reductions associated with a booster session), we did not find another study in the literature that explored the effects of saturation of an intervention across the middle school grades versus one grade receiving a DV/H intervention.

In 2011, the same team started a third experiment (referred to hereafter as the "NYC-2" study). This new study extends the earlier work in Cleveland and NYC-1 by: Expanding the study to include 8th grade as well as 6th and 7th grade students, including the use of a few lessons from Safe Dates (Foshee et al., 1998) for 8th graders, testing revised grade-specific DV/H interventions for middle schools (our earlier work used non-grade differentiated classroom lessons), and having a longer follow-up period of 12 months instead of the 6 months follow-up of the NYC-1 research. Overall, this study helps address several gaps in the literature regarding how often to intervene, and how broadly to intervene across grade levels in order to produce a safe environment.

Project Purpose, Goals and Objectives

The long-term goal of this study is to help reduce and prevent DV/H among middle school students by employing rigorous methods to evaluate the evidence-based *Shifting Boundaries* (SB) program. The purpose of this study was to provide high-quality scientific evidence concerning the effectiveness of a multi-level longitudinal approach to DV/H prevention programming for public middle school students from New York City (NYC). Our first objective was to assess whether and how much of a difference it makes when all three grades in a middle school receive a set of DV/H classroom and building-based interventions compared to when only two of the three grades receive it or only one grade receives it. To address this question, we built on our earlier research to assess the impact of *saturating* a middle school environment with information and behavioral strategies tailored to each grade level—a departure from our earlier work that looked at only the 6th and the 7th grades—acknowledging the integrated social environment of a middle school that may contribute to ongoing dating violence. Our second objective was to examine the impact of multiple *doses* of grade-differentiated curricula and

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

whether additional dosages over two years of SB leads to greater reductions in DV/H than single dosages in one year. Overall, this study helps address gaps in the literature regarding how broadly to intervene across grade levels and how often to intervene in order to produce a safe school environment.

Methods

This report provides a detailed account of the NYC-2 results of an experimental evaluation that used a randomized controlled trial (RCT) of a DV/H prevention program — the "Shifting Boundaries Program"—for sixth, seventh and eighth grade students in New York City. We randomly assigned 23 public middle schools in New York City to one of four treatment conditions of this school-based intervention. The unit of assignment and unit of analysis were schools. With this type of design (students nested within schools), we added a statistical correction in our models to provide for robust clustered standard errors. Our study employed a multi-level, experimental, longitudinal design. The study included 23 middle schools, four treatment groups and three waves of student self-report surveys (baseline, 6-month follow-up and 12 month follow-up). We examined schools that we provided varying levels of dosages of SB (inclusive of SBC and SBS interventions):

- **Group 1**: schools assigned here received SB in one school year for 6th graders only,
- **Group 2**: schools received SB in one school year for their 6th and 7th grades,
- **Group 3**: schools received SB in one school year for their 6th, 7th and 8th grades,
- **Group 4**: schools received SB over two school years first in 6th grade and the same group receives it in 7th grade the following school year.

The intervention we tested had two main components. First, we had an SB classroom curriculum (SBC), covering the consequences for perpetrators of DV/H, laws and penalties for DV/H, and respectful relationships. Second, we had an SB school (building-level) (SBS) component which included the use of school-based restraining orders, higher levels of faculty and security presence in areas identified through student mapping of safe/unsafe "hot spots," and the use of posters to increase awareness and reporting of DV/H. We examined (1) the effects of saturating a school environment by providing the SB intervention to all three middle school grades compared to only two grades or one grade and (2) the effects of two dosages of SB across two years compared to one dosage of the SB intervention across one year.

We had to modify our planned design for this experiment due to low levels of participation among the New York City middle schools. While we planned to have a no-treatment control group that proved not feasible in this context, we decided to maximize our use of the schools to address our main research question of the comparative effectiveness of different levels of SB treatment. Also, we had a no treatment control group with our earlier NYC-1 experiment conducted only a couple of years before the current study and it already addressed the question of treatment efficacy (treatment versus no-treatment). Next, we hoped to have our 6th grade longitudinal group receive three years of treatment. However, too few of the schools were willing to continue participation beyond two years of treatment due to competing academic demands within the schools and staffing deficits leading to a resistance to applying limited resources to DV/H prevention. Similarly, we had hoped to have up to a 24-month

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

follow-up survey but that plan had to be abandoned given the lack of willingness of schools to continue participation. The schools agreed to one-year of follow-up surveys. Nevertheless, the one-year follow-up still represented a longer follow-up period than the earlier research on SB that only had a six-month follow-up period (Taylor et al., 2008; Taylor et al., 2013a; Taylor, 2010). Despite these problems, we continued the project and in the end we believe some important findings emerged and we learned some important lessons about navigating a field experiment through multiple administrative issues.

Description of Participants

Participating students ranged in age from 10 to 15, with a nearly 50% split between boys and girls. Our sample was 26% Hispanic, 37% African American, 16% Asian, 14% White and 7% "other." Over 40% of the sample had prior experience with a violence prevention educational program. Nearly half reported at least one experience of being in a dating relationship. At baseline, about one in five respondents reported having ever been the victim of any physical dating violence, with a similar number reporting perpetrating any physical dating violence. One in ten respondents reported having been the victim of any sexual dating violence ever (6.4% for perpetration of this act). Almost 60% of the sample reported having ever been the victim of any physical peer violence at some point in time (45% perpetration), and 18.1% were ever the victim of sexual peer violence (8% perpetration). Also, 49% reported experiencing sexual harassment (SH) at some point in time (23% perpetration).

Results

We conducted our analyses using Stata 12.0 statistical software with a robust variance estimator. Given our substantive interest in the individual data, and our need to only control for the clustering of the school-level data, the use of a robust variance estimator to address the clustered nature of our data and produce unbiased estimates was adopted (Rogers, 1993; Williams, 2000). We present outcome models for our Wave 2 (6-months post treatment) and Wave 3 (about 12 months post treatment) data for sexual harassment (8 models), and violent behavior (16 models for peers and for dating relationships [total 32 models]).

6-month follow-up outcome data: Our six month follow-up data allows us to address the saturation question of how much of a difference it makes when only the 6th grade in a middle school receives SB compared to higher levels of saturation when the 6th and 7th grade in a middle school receive SB or when all three grades in a middle school receive SB. For these analyses we combined our 6th grade only group (n= 3 schools) with our 6th grade longitudinal group (n= 8 schools) because at the sixth month follow-up mark the 6th grade longitudinal group had only received the 6th grade intervention and was functionally the equivalent of the 6th grade only group. The combined group had 11 schools and 631 students. Our next group is made up schools assigned to receive the 6th and 7th grade SB intervention (n= 3 schools and 271 students). The final group is made up schools assigned to receive our 6th, 7th and 8th grade SB intervention (n= 9 schools and 862 students).

Our first set of models covered the <u>prevalence</u> of peer violence and sexual harassment at the six-month post intervention mark (including victimization and perpetration). None of the

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

treatment comparisons between the reference category of the 6th grade-only group to the 6th and 7th grade group and the 6th, 7th and 8th grade group were statistically significant, suggesting that the treatment effects were equal. Our second set of models covered the frequency of peer violence and sexual harassment at the six-month post intervention mark. We had one statistically significant effect (i.e., $p < .05$). The 6th & 7th grade dosage was associated with less sexual harassment victimization frequency compared to the reference category of 6th grade only (Regression Coefficient = -0.18, p= 0.01). The standardized effect size for this finding was a Cohen's D of 0.20 which is equivalent to a small effect just below a medium sized effect (Cohen, 1988). We also had two borderline cases ($p<.10$). The 6th & 7th grade group was associated with less peer physical victimization frequency compared to the reference category of 6th grade only (Regression Coefficient = -0.12, p= 0.08). The 6th & 7th grade group was associated with less peer sexual perpetration frequency compared to the reference category of 6th grade only (Regression Coefficient = -0.06, p= 0.09). However, none of the other treatment comparisons between the reference category of the 6th grade-only group to the 6th and 7th grade group and the 6th, 7th and 8th grade group were statistically significant.

Our third set of models covered the prevalence of physical dating violence and sexual dating violence at the six-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the 6th and 7th grade group and the 6th, 7th and 8th grade group were statistically significant, suggesting that the treatment effects were equal. Our fourth set of models covered the frequency of physical dating violence and sexual dating violence at the six-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the 6th and 7th grade group and the 6th, 7th and 8th grade group were statistically significant, suggesting that the treatment effects were equal.

12-month follow-up outcome data: Our twelve-month follow-up data allows us to address the question of whether additional dosages of SB leads to greater reductions in DV/H than single dosages (6th grade longitudinal versus 6th grade only groups) and assess our first saturation-level question on the relative value of multiple grades receiving SB compared to one grade at the 12-month follow-up mark. For these analyses we do not combine our 6th grade only group (n= 3 schools with 128 students) with our 6th grade longitudinal group (n= 4 schools with 234 students) because at the 12-month follow-up mark the 6th grade longitudinal group has received the 6th grade intervention when they were in sixth grade and the 7th grade intervention when they were in seventh grade. As the name indicates, the 6th grade only group only receives the intervention in 6th grade and not again in 7th grade. Our third group is made up schools assigned to receive the 6th and 7th grade SB intervention or the 6th, 7th and 8th grade SB intervention (n= 6 schools and 452 students for both sets of schools).

Our first set of 12-month outcome models covered the prevalence of peer violence and sexual harassment at the twelve-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the combined 6th and 7th grade group with the 6th, 7th and 8th grade group

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

or the 6th grade longitudinal group were statistically significant, suggesting that the treatment effects were equal.

Our second set of 12-month outcome models covered the frequency of peer violence and sexual harassment at the six-month post intervention mark. We had two statistically significant effects (i.e., $p < .05$). The 6th & 7th grade intervention or 6th, 7th & 8th grade interventions were associated unexpectedly with more peer sexual violence perpetration compared to the 6th grade only intervention (Regression coefficient = 0.04, p= 0.04). The standardized effect size for this finding was a Cohen's D of 0.23 which is equivalent to a small effect (Cohen, 1988). Our next statistically significant finding was that the 6th & 7th grade intervention or 6th, 7th & 8th grade intervention were associated with (as we anticipated) less sexual harassment victimization compared to the 6th grade only intervention (Regression coefficient = -0.22, p= 0.03). The standardized effect size for this finding was a Cohen's D of 0.26 which is equivalent to a small effect (Cohen, 1988). We also had one borderline statistically significant finding ($p<.10$). Our 6th grade longitudinal group that received two dosages of the SB intervention (one in 6th and one in 7th grade) was associated with (as we anticipated) less sexual harassment victimization compared to the 6th grade only intervention (Regression coefficient = -0.18, p= 0.08). The standardized effect size for this finding was a Cohen's D of 0.22 or a small effect (Cohen, 1988). None of the other treatment comparisons between the reference category of the 6th grade-only group to the other groups were statistically significant.

Our third set of 12-month outcome models covered the prevalence of physical dating violence and sexual dating violence at the twelve-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the combined 6th and 7th grade group with the 6th, 7th and 8th grade group or the 6th grade longitudinal group were statistically significant, suggesting that the treatment effects were equal.

Our fourth set of 12-month outcome models covered the frequency of physical dating violence and sexual dating violence at the twelve-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the combined 6th and 7th grade group with the 6th, 7th and 8th grade group or the 6th grade longitudinal group were statistically significant, suggesting that the treatment effects were equal.

Key Study Limitations

The general limitations of self-reports are applicable to this study. For example, students may not be able to recall the timing of a violent act or may have deliberately under-reported certain behavior (Jackson, Cram, & Seymour, 2000), or may have exaggerated certain behavior. Despite these potential problems, which likely were balanced across our comparison groups, self-report surveys have become an accepted modality of collecting violence data. Next, because of concerns raised by school personnel on the sensitivity of such questions for a middle school population, we were limited in how we could measure sexual victimization to two main

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

items. Also, our study was limited to six and twelve month follow-up periods and it is unclear whether our findings would change over a longer follow-up period.

There are several threats to the validity of our experiment. Based on our analyses, we believe that our experimental design created comparable conditions to assess outcome differences in our comparison groups. While we found some differences between the comparison conditions prior to the experiment (during the baseline period), most of these differences (while statistically significant) were not very large differences. For the most part, the four study groups/conditions were similar on the majority of our measures leaving the only major differences across the groups their assigned intervention condition. Additionally, random assignment procedures were followed closely (no "overrides"). Despite schools dropping out of the study, the schools that stayed in the study adhered to their assigned treatment. Finally, we included the variables where there were pre-treatment differences into our outcome models as covariates to remove any potential biases these small imbalances might have presented for the interpretation of our results.

Another major concern in our study was whether attrition in our study created any pattern of bias that would interfere with our ability to draw unequivocal inferences from our study. Overall, we did not observe much by way of patterns in our study for the schools that continued on to complete the follow-up survey waves and those schools that dropped out after doing only a baseline survey. We observed few differences between the dropout schools and the completer schools on a variety of background factors and violence measures. Where there were some differences, we addressed this in our statistical modeling.

Implications and Conclusion

Despite the difficult path it took to complete the NYC-2 RCT, and the associated changes we had to make to the study (dropping our control group, reducing the dosage testing to two instead of three dosages, and reducing the follow-up period to 1 year instead of potentially 3 years), we still believe some important new knowledge emerged from the NYC-2 study. In addition to lessons learned about how to navigate a field experiment through difficult administrative and resource barriers, we learned after NYC-2 that for the most part additional dosages and saturation do not alter the findings for most of our outcome measures compared to just implementing the SB program with just 6th grade students. Providing the SB treatment once in 6th grade works as well (in terms of DV/H levels) as applying it once per year for two years with the same group (in 6th and 7th grades). Likewise, implementing SB with only one grade level in a middle school does just as well in terms of peer violence and dating violence outcomes as a more saturated process of treating multiple grades in the school. However, we did find that additional saturation beyond one grade is associated with reductions in sexual harassment victimization at the 6 and 12-month follow-up period. Considering our NYC-1 and NYC-2 results together, we believe there is empirical justification for implementing the school-wide component SBS across the entire school environment. That is, we feel the data support implementing the SB program for at least the 6th grade students but given the nature of the SBS intervention that can be extended to the whole school with little extra cost.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Introduction

Teen dating violence and sexual harassment (DV/H) (including "interpersonal" or "gender" violence)[2] have been recognized as serious and persistent problems in the public health and violence prevention fields (Centers for Disease Control and Prevention, 2012a, 2014; Jouriles, Garrido, et al., 2009; Mulford & Giordano, 2008a; National Institute of Justice, 2011; Shanklin et al., 2007). Local and regional studies reveal that approximately 50-60% of teens are victims of teen dating violence (TDV) (Foshee, 1996; Hickman et al., 2004; Jouriles, Platt, et al., 2009; Malik et al., 1997; O'Keefe, 1997). Between 10-15% of adolescents are victims of physical TDV (Manganello, 2008). As of 2014, drawing on the national Survey of Teen Relationships and Intimate Violence (STRiV), two-thirds of U.S. adolescents in a current or past-year dating relationship experienced psychological victimization, and nearly as many admitted to perpetrating the same (Taylor & Mumford, 2014 e-pub ahead of print). One in five reported physical and/or sexual victimization, while one in eight adolescents reported perpetrating physical and/or sexual violence in a dating relationship (Taylor & Mumford, 2014 e-pub ahead of print).

In addition to the risk of physical injury, teen dating violence is associated with significantly poorer mental and physical health (Howard et al., 2007b), suicide attempts (Chiodo et al., 2012a), risky sexual behavior (Hanson, 2010), teen/unwanted pregnancies (Silverman, Raj, & Clements, 2004), and unhealthy weight control (Ackard & Neumark-Sztainer, 2002). Teen dating violence is also a risk factor for adult intimate partner violence (Berkowitz, 2010; Gomez, 2010), with as much as half of intimate violence persisting into adulthood (Halpern, Spriggs, Martin, & Kupper, 2009).

Among adolescents, along with increases in dating behavior, the onset of puberty is accompanied by increases in sexual harassment (SH) (Pepler et al., 2006), which may occur within or outside dating or peer relationships. The 2011 AAUW study found in a nationally representative sample of youth (12 to 18 years old) that about half of their sample reported being a past year victim of SH (Hill & Kearl, 2011). Even higher rates of SH were reported in the 2001 AAUW national study that found that 83% of females experience SH by male peers while male students also reveal high levels (60-79%) of SH in school (AAUW 2001). STRiV researchers found no statistically significant gender differences in SH victimization rates but did find higher rates for boys perpetrating SH compared to girls (Taylor & Mumford, 2014 e-pub ahead of print). School-based SH interferes with the educational experience and constitutionally granted right to attend school in an environment that is free from sex discrimination and harassment ("Davis v. Monroe County Board of Education," 1999; "Franklin v. Gwinnett County (GA) Public

[2] In this report, we use the term *dating violence and harassment* (DV/H) to represent physical, emotional, or sexual abuse within a dating relationship, the definition that CDC uses for teen dating violence (TDV) (Centers for Disease Control and Prevention, Accessed 2/11/11). More broadly, this problem has also been referred to as gendered adolescent interpersonal aggression (GAIA) (Smith et al., 2009) Where cited studies used the term TDV, we also follow the language of the original research. We also will use the term DV/H to cover youth sexual harassment (SH). We follow the American Association of University Women's (AAUW) definition of SH. SH is unwelcome conduct of a sexual nature in person or through electronic means, which can include unwelcome sexual advances, requests for sexual favors, or other verbal, nonverbal, or physical conduct of a sexual nature (Hill & Kearl, 2011).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Schools," 1992; U.S. Department of Education (2001); US Department of Education (1997)). Yet historically SH has often been tolerated and even normalized by school administrators and students alike (American Association of University Women, 1993, 2001; Stein, 1995, 1999).

Despite the grim news on the vast prevalence and far reaching impact of DV/H, over the past decade a number of interventions have been developed to prevent DV/H and/or lower its reoccurrence. Rigorous research has been conducted on the effectiveness of DV/H prevention programs (Foshee & Reyes, 2009; Jaycox et al., 2006; Taylor, 2010; Wolfe et al., 2009), and a number of these studies have shown positive results. However, these studies are few and generally address only 8th and/or 9th grade or older students (Foshee et al., 1998; Lisa H. Jaycox et al., 2006; Wolfe et al., 2009). Only a few addressed 6th and 7th grade students (Peskin et al., 2014; Taylor, 2010).

This report provides a detailed account of the results of an experimental evaluation that used a randomized controlled trial (RCT) of a DV/H prevention program — the "Shifting Boundaries Program"—for sixth, seventh and eighth grade students in New York City. We randomly assigned 23 public middle schools in New York City to one of four treatment conditions of this school-based intervention. The project includes a baseline and two follow-up surveys with 6th, 7th and 8th grade students to assess short-term environmental impacts; intermediate changes and long term impact on rates of TDV. The classroom intervention was delivered through a multi-session curriculum that emphasized the consequences for perpetrators of DV/H, state laws and penalties for DV/H, and respectful relationships. The school (building-level) intervention included the development and use of temporary school-based restraining orders, higher levels of faculty and security presence in areas identified through student mapping of safe/unsafe "hot spots," and the use of posters to increase awareness and reporting of DV/H to school personnel.

Through our RCT, we are investigating two key questions related to: (1) the effects of saturating a school environment by providing the SB intervention to all three middle school grades compared to only two grades or one grade and (2) the effects of two dosages of the SB intervention across two years compared to one dosage of the SB intervention across one year. Overall, this study helps address gaps in the literature regarding how broadly to intervene across grade levels and how often to intervene in order to produce a safe school environment. Our study was designed to yield policy relevant data that could help increase the capacity of schools to prevent DV/H. In the long-term, we hope that the results from this study could be used to help prevent DV/H and other forms of violence and harassment. To achieve this goal we employed rigorous methods to provide clear results on the effectiveness of strategies for altering the violence-supportive attitudes and norms of youth.

The sections that follow include a review of the extent DV/H research literature, a detailed presentation of the research methods used in our study, a summary of our theoretical framework for our interventions and the study hypotheses, our study results, a discussion of our study results and some concluding comments.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Literature Review

To follow is a review of the DV/H research literature on the scope of DV/H to provide a sense of the extent of this problem that the SB program is designed to address. Next, we cover the nature of DV/H, including covering relevant material on the onset of DV/H and developmental pathways that lead to DV/H. This section is followed by material on the extensive consequences associated with DV/H and finally we cover prior scientific studies evaluating school-based violence prevention programs.

Dating Violence and Sexual Harassment: Scope of the Problem

The focus of our intervention is the problem of youth DV/H which by a number of accounts is a significant public health problem. First, TDV can take the form of physical abuse, emotional/verbal/psychological abuse, or sexual abuse (Offenhauer & Buchalter, 2011). The situational venue may be in person or via electronic means, in both public and private spaces, between current or past dating partners (Centers for Disease Control and Prevention, 2012b). Local and regional studies reveal that approximately 50-60% of teens are victims of TDV (Foshee, 1996; Hickman et al., 2004; Jouriles, Platt, et al., 2009; Malik et al., 1997; O'Keefe, 1997). Between 10-15% of adolescents are victims of physical adolescent relationship abuse (ARA) (Manganello, 2008). As of 2014, drawing on the national Survey of Teen Relationships and Intimate Violence (STRiV), two-thirds of U.S. adolescents in a current or past-year dating relationship experienced psychological victimization, and nearly as many admitted to perpetrating the same. One in five reported physical and/or sexual victimization, while one in eight adolescents reported perpetrating physical and/or sexual violence in a dating relationship (Taylor & Mumford, 2014 e-pub ahead of print). Compared to youth ages 15-18, those 12-14 years old in STRiV reported lower rates of psychological and sexual ARA victimization; similarly, we found lower ARA perpetration rates for those 12-14 (Taylor & Mumford, 2014 e-pub ahead of print). While other researchers, with a few exceptions (Arriaga & Foshee, 2004; Molidor & Tolman, 1998), have found that girls experience ARA victimization at higher rates than boys (Bergman, 1992; Hamby, Finkelhor, & Turner, 2012; Wolitzky-Taylor et al., 2008a), the STRiV research team found no gender differences for ARA victimization, but found that girls perpetrated more physical ARA (based on the Conflict in Adolescent Dating Relationships Inventory – CADRI (Wolfe et al., 2001)) than boys (Taylor & Mumford, 2014 e-pub ahead of print)

The existence of peer-to-peer sexual harassment in K-12 schools has been well documented for decades (American Association of University Women, 1993, 2001; Stein, 1981; Stein, 1995, 1999; Stein, Marshall, & Tropp, 1993; Straus, 1988). The American Association of University Women's national survey of sexual harassment in schools in 2001 found 83% of girls and 79% of boys indicating that they had been sexually harassed (American Association of University Women, 2001). In the 2001 study, thirty percent of girls and 24% of boys reported that they were sexually harassed often (American Association of University Women, 2001) and 60–79% of boys reported being verbally harassed (American Association of University Women, 1993, 2001; Tolman, Spencer, Rosen-Reynoso, & Porche, 2003). The 2011 AAUW study found in a nationally representative sample of youth (12 to 18 years old) lower rates than the 2001 study

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

but still found that about half of their sample reported being a past year victim of SH (Hill & Kearl, 2011).

The Nature of DV/H

The nature of DV/H has implications for interventions designed to prevent this problem. In this section, we explore the nature of the problem of DV/H including its onset and developmental pathways and the consequences associated with DV/H.

Onset and developmental pathways. While the majority of research in this realm focuses on students in grades 8-12 (Foshee et al., 1998; Foshee, Bauman, et al., 2005; Foshee et al., 2000; Foshee, Bauman, Linder, Rice, & Wilcher, 2007; Foshee, Linder, MacDougall, & Bangdiwala, 2001; Foshee et al., 1996a, 1996b; Foshee & Matthew, 2007; Jaycox et al., 2006), it is still informative for our research.

Although formal dating is limited among younger adolescents, early gendered conflicts have been documented (Noonan & Charles, 2009b). Sexual harassment tends to increase throughout middle school (Bentley, Galliher, & Ferguson, 2007; Manganello, 2008; McMaster, Connolly, Pepler, & Craig, 2002; Pellegrini, 2001), indicating that middle schools can serve as a training ground for DV/H (Stein, 1995) and indicating a window of opportunity for early intervention (Mulford & Giordano, 2008b; Noonan & Charles, 2009a). Socialization experiences and early childhood exposure to violence are predictive factors for the perpetration of sexual violence (Nagayama Hall & Barongan, 1997). The research evidence to-date suggests that youth can become a victim of DV/H and sexual harassment as early as 6th grade (Callahan, Tolman, & Saunders, 2003; Eaton et al., 2010; O'Keefe, 1997), suggesting that prevention programming needs to start as early as middle school (American Association of University Women, 2001; Basile, Espelage, Rivers, McMahon, & Simon, 2009; Burcky, Reuterman, & Kopsky, 1988; Espelage & Holt, 2007; Foshee et al., 1998; McMaster et al., 2002; Meyer & Stein, 2004; Schewe, 2000a, 2002; Taylor et al., 2008; Taylor, Mumford, Hair, Stein, & Yu, 2010; Tolman et al., 2003; Wolitzky-Taylor et al., 2008b). For example, in a sample of 6th and 7th grade students in the Cleveland area 19% reported being sexually victimized by a school peer (Taylor et al., 2008; Taylor, 2010) and among a sample of 7th grade students who had begun dating , one-third reported TDV (Sears, Byers, & Price, 2007).

Consequences associated with DV/H. The pressing nature of developing interventions to address DV/H is supported by the appreciable consequences associated with this problem. While there is limited rigorous longitudinal data on youth DV/H consequences, the best available data shows that victims of SH have significantly poorer mental and physical health, more trauma symptoms, and greater school avoidance than those not sexually harassed (AAUW Educational Foundation, 1993, 2001; Gruber & Fineran, 2008; Hill & Kearl, 2011; Larkin, 1994). Youth experiencing SH were more likely to report depression, loss of appetite, nightmares or disturbed sleep, low self-esteem and feelings of being sad, afraid, scared, or embarrassed (Hand & Sanchez, 2000; Lee, Croninger, Linn, & Chen, 1996).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Also, TDV is associated with a range of poor outcomes, including significantly poorer mental and physical health for teens (Howard et al., 2007b; Molidor & Tolman, 1998), suicide attempts (Chiodo et al., 2012a), and higher risks of depression and anxiety (Banyard & Cross, 2008; Callahan et al., 2003; Exner-Cortens, Eckenrode, & Rothman, 2013; Holt & Espelage, 2005; Howard & Wang, 2003a; Howard & Wang, 2003b; McDonald, Graham, & Martin, 2010) and suicide attempts (Chiodo et al., 2012a). TDV is also associated with substance use (Ackard, Eisenberg, & Neumark-Sztainer, 2007; Banyard & Cross, 2008; Chiodo, Wolfe, Crooks, Hughes, & Jaffe, 2009; Coker et al., 2000; DuRant et al., 2000; Fineran & Bolen, 2006; Hanson, 2010; Holt & Espelage, 2005; Howard & Wang, 2003b; Kreiter et al., 1999; Mendelson, Turner, & Tandon, 2010; Roberts & Klein, 2003; Roberts, Klein, & Fisher, 2003; Silverman, Raj, Mucci, & Hathaway, 2001), risky sexual behavior (Champion, Foley, Sigmon-Smith, Sutfin, & DuRant, 2008; Chiodo et al., 2009; Hanson, 2010; Holt & Espelage, 2005; Howard & Wang, 2003b; Silverman et al., 2001), unwanted fertility outcomes (Champion et al., 2008; Chiodo et al., 2009; Mendelson et al., 2010; Shrier, Pierce, Emans, & DuRant, 1998; Silverman et al., 2004; Silverman et al., 2001), unhealthy weight control (Ackard & Neumark-Sztainer, 2002; Silverman et al., 2001), other trauma symptoms (Howard et al., 2007a, 2007b; Molidor & Tolman, 1998), and adult intimate partner violence (Berkowitz, 2010; Gomez, 2010).

While there has been a considerable amount of research, as documented above, on the consequences of ARA, few of these studies have used longitudinal designs to disentangle these consequences from other spurious factors and whether these behaviors are a cause or consequence of ARA. In longitudinal research on the association between TDV victimization and later adverse outcomes (Ackard et al., 2007; Exner-Cortens et al., 2013; Exner-Cortens D., 2012; Gomez, 2011; Halpern et al., 2009; Roberts et al., 2003; Spriggs, Halpern, & Martin, 2009; Teitelman, Ratcliffe, Dichter, & Sullivan, 2008; Van Dulmen et al., 2012), only a small number have investigated outcomes other than risk for revictimization. In a study based on Waves 1 (1994–1995), 2 (1996), and 3 (2001–2002) of the National Longitudinal Study of Adolescent Health (ADHEALTH), a nationally representative sample of US high schools and middle schools, compared with participants reporting no teen dating violence victimization at Wave 2, female participants experiencing victimization reported increased heavy episodic drinking, depressive symptomatology, suicidal ideation, smoking, and IPV victimization at Wave 3, whereas male participants experiencing victimization reported increased antisocial behaviors, suicidal ideation, marijuana use, and IPV victimization at Wave 3, controlling for sociodemographic variables, child maltreatment, and pubertal status (Exner-Cortens et al., 2013).

In another analysis of ADHEALTH longitudinal data, Gomez developed a series of regression models stratified by gender, and found that child abuse and adolescent dating violence are significant predictors of IPV victimization and perpetration for both men and women (Gomez, 2011). Experiencing violence during childhood and/or adolescence is highly predictive of IPV in young adulthood (Gomez, 2011). In Van Dulmen and colleagues' analysis of ADHEALTH longitudinal data, they found that suicidality was associated with IPV victimization prospectively, but IPV victimization did not predict suicidality prospectively (Van Dulmen et al., 2012).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

In one of the few studies to not rely on ADHEALTH, Ackard and colleagues looked at the effects of physical and sexual ARA on adverse health outcomes 5 years post-victimization in a sample of 1,516 Minnesota older adolescents (Ackard et al., 2007). In analyses adjusted for wave 1, in female adolescents, ARA was significantly associated with smoking cigarettes, marijuana use, and high depressive symptoms and marginally associated with suicide attempts (Ackard et al., 2007). In boys, ARA was significantly associated with smoking cigarettes and marginally associated with binge-eating and suicidal ideation (Ackard et al., 2007). ARA was significantly associated with an overall high-risk profile (presence >3 health outcomes) for both girls and boys; results remained significant in female adolescents after adjusting for wave 1 (Ackard et al., 2007).

Using five waves of structured interview data from the Toledo Adolescent Relationships Study (TARS), Copp and Johnson found that while young women generally report higher levels of depression, IPV victimization is similarly linked to variability in men's and women's self-reports of depressive symptoms (Copp & Johnson, 2015). They also found that reports of IPV (both victimization and perpetration) were associated with declines in self-rated health among men and women (Copp & Johnson, 2015). A more focused analysis relying on waves 4 and 5 of the data revealed that perpetration and victimization were associated with changes in levels of anxiety across these two points in time among both male and female respondents (Copp & Johnson, 2015).

School-based Violence Prevention Programs

While prevention efforts about other forms of general youth violence enjoy widespread support, programs to prevent adolescent DV/H have emerged more slowly (Wekerle & Wolfe, 1999). Based on their review of the research on dating violence prevention programs, Cornelius and Resseguie (2006) note that most prevention evaluations have documented at least a short-term positive change in knowledge and/or attitudes related to youth DV/H prevention (AveryLeaf, Cascardi, Oleary, & Cano, 1997; Foshee, 1996; Foshee et al., 1998; Foshee et al., 2004b; Foshee et al., 2000; Foshee, Benefield, Ennett, Bauman, & Suchindran, 2004; Foshee et al., 1996a; Jaffe, Sudermann, Reitzel, & Killips, 1992; Lavoie, Vezina, Piche, & Boivin, 1995; MacGowan, 1997; Ward, 2002), while others show longer-term positive program effects (Foshee et al., 2004b; Foshee, Bauman, et al., 2005; Foshee, Benefield, et al., 2004; Wolfe et al., 2009).

However, much of this research was <u>not</u> done with rigorous designs such as RCTs or quasi-experiments (Centers for Disease Control and Prevention, 1998; Chalk, King, & Eds., 1998; Meyer & Stein, 2004; Ward, 2002), and most studies are of high school students (AveryLeaf et al., 1997; Centers for Disease Control and Prevention, 1998; Chalk et al., 1998; Cornelius & Resseguie, 2006; Foshee et al., 1998; Foshee, Bauman, et al., 2005; Foshee et al., 2000; Foshee & Reyes, 2009; Hickman et al., 2004; Jaycox et al., 2006; Jones, 1991; Lavoie et al., 1995; Meyer & Stein, 2004; Pacifici, Stoolmiller, & Nelson, 2001; Whitaker et al., 2006; Wolfe et al., 2009; Wolfe et al., 2003). While relatively more commonplace in recent years in criminology (Braga, Welsh, Papachristos, Schnell, & Grossman, 2014), conducting field RCTs are still challenging and become even more complicated when the study involves youth (as young as 10 years old) and

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

bureaucracies as complex as public middle schools. For that reason perhaps many of the existing TDV experiments (n=4 studies) have been done with high school students (Jaycox et al., 2006; McMahon & Banyard, 2011; Silverschanz, Cortina, Konik, & Magley, 2008; Wolfe et al., 2009) or high school and 8th grade (n= 1 study) middle school students (Foshee et al., 1998; Foshee, Bauman, et al., 2004). Only two other TDV experiment, outside the Shifting Boundary experiments (Taylor, Stein, & Burden, 2010a; Taylor, Stein, & Burden, 2010b; Taylor, Stein, Mumford, & Woods, 2013b) , have been done exclusively with middle school students: one in Texas (Peskin et al., 2014) and one in the Bronx, NY (Cissner & Ayoub, 2014).

Safe Dates, a U.S.-based program for 8th and 9th grade students designed by Foshee et al. (Foshee et al., 1998; Foshee, Bauman, et al., 2004; Foshee, Bauman, et al., 2005; Foshee et al., 2000; Foshee et al., 1996), has experimentally shown a reduction in long-term physical dating violence. While *Safe Dates* is now relatively well known and implemented, the research was based on a small rural sample (n=14 schools). In another of the more rigorously evaluated interventions (the *Fourth R: Skills for Youth Relationships* program delivered to Canadian 9th graders), Wolfe and colleagues (2009) found that after 21 sessions the program for the 9th grade Canadian students was able to reduce physical dating violence in the intervention group as compared to the control group up to 2.5 years post treatment. Based on this experimental research, Wolfe and colleagues (2009) called for interventions with younger students.

In a group RCT with 10 middle schools in southeast Texas (n= 766 predominantly ethnic-minority students), Peskin and colleagues found that those receiving the TDV prevention program called *It's Your Game…Keep It Real* (IYG) experienced less physical TDV victimization and emotional TDV victimization and perpetrated less emotional TDV than the control group (Peskin et al., 2014). IYG did not have an effect on physical dating violence perpetration. Overall, IYG significantly reduced three of four dating violence outcomes among ethnic-minority middle school youth.

In another evaluation of the *Fourth R: Skills for Youth Relationships* program, incoming seventh-grade students in ten Bronx middle schools were assigned to class sections, which were then randomly assigned to receive the Fourth R or a standard seventh-grade curriculum (the control group) during the 2011-2012 academic year (Cissner & Ayoub, 2014). The main effects analyses showed <u>no impact</u> of the Fourth R curriculum on dating violence, peer violence/bullying, or drug and alcohol use (Cissner & Ayoub, 2014), perhaps attributable to the low statistical power present in the study.

Finally, there have been other developments in the field. In FY 2009, Congress began providing the CDC with funding to rigorously address the problem of TDV. With this funding, CDC developed a comprehensive TDV prevention initiative called "*Dating Matters: Strategies to Promote Healthy Teen Relationships*." The *Dating Matters Initiative* serves as a demonstration project for preventing youth interpersonal violence within families and among acquaintances. Further, the prevention program targets middle school-aged individuals with varying curricula for 6th, 7th and 8th graders. The CDC is evaluating the *Dating Matters Initiative* with an RCT design and the evaluation is ongoing.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

While not with middle schools, another prominent program "Coaching Boys into Men" (CBIM) found some encouraging results for a high school population (Miller, Tancredi, McCauley, Decker, Virata, Anderson, O'Connor, et al., 2013; Miller et al., 2012). This cluster RCT explored the effectiveness of a DV perpetration prevention program targeting coaches and high school male athletes (Miller et al., 2012). Sixteen high schools were randomly assigned to either receive CBIM or a control condition of 'coaching as usual' (Miller et al., 2012). CBIM consists of a 60-minute training for coaches led by a trained violence prevention advocate to introduce the Coaches Kit (available at http://coachescorner.org), which provides strategies for opening conversation about violence against women with athletes (Miller et al., 2012). Eleven "Training Cards" guide coaches to lead brief (10 –15 min) weekly discussions with athletes about respect and DV prevention through the sports season (Miller et al., 2012). Compared to control subjects, the program had a statistically significant effect on positive intentions to intervene and higher levels of positive bystander intervention (Miller et al., 2012). The RCT did not yield positive changes in ARA perpetration behavior (Miller et al., 2012). The same research team also followed up with this group of students 12 months after they had received the CBIM intervention (Miller, Tancredi, McCauley, Decker, Virata, Anderson, O'Connor, et al., 2013) and found reductions in reported ARA perpetration in the past three months for the treatment group compared to the control group, and the treatment group was less likely to go along with peers' abusive behaviors. However, there was no long-term effect of CBIM on gender-equitable attitudes, intervening as bystanders, or recognition of abusive behaviors (Miller, Tancredi, McCauley, Decker, Virata, Anderson, O'Connor, et al., 2013).

Related to school-based curriculum-based ARA prevention programs, are bystander intervention programs for college campuses which have been recognized as a promising prevention strategy by the 2014 White House Task Force to Protect Students from Sexual Assault (White House Task Force to Protect Students from Sexual Assault, 2014). Bystander programs are innovative by engaging participants as possible witnesses to violence rather than potential victims or perpetrators and providing skills to reduce the risk for violence for themselves and those in their peer group. Results from rigorous research indicate that bystander programming can increase bystander intentions (Potter & Moynihan, 2011), behaviors (Coker et al., 2011) and reduce interpersonal violence among college students (Coker et al., 2014). More recently researchers have begun to apply bystander programs to high schools. Using a cluster RCT, Eckstein and Edwards are testing the efficacy of an adapted version of Bringing in the Bystander® (BITB), originally created for college-age students, with high school youth (Eckstein, 2014). The team is randomly assigning 26 New England high schools to receive either the Bringing in the Bystander—High School Curriculum (BITB-HSC) or a control condition (Eckstein, 2014).

Prior Shifting Boundaries (SB) Experiments

Our current study is an outgrowth of a program of research on the Shifting Boundaries intervention. In 2005, a research team led by Taylor and Stein (Taylor & Stein, Ohio 2005-2007) conducted one of the first experimental evaluations of a primary prevention program addressing DV/H for sixth and seventh grade students. The study was conducted in suburban

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

middle schools bordering Cleveland, funded by the U.S. Department of Justice (DOJ). This evaluation was the first to work with students as young as 10 to 12 years old in sixth and seventh grades on DV/H. This earlier research confirmed the importance of reaching middle school students with prevention programming (Taylor et al., 2008; Taylor, 2010). Our team developed two 5-lesson curricula to address DV/H. As discussed below, classrooms were randomly assigned to one of three conditions. Our first treatment was an interaction-based curriculum (described below). Our second treatment was a law and justice curriculum and we had a control group who did not receive either treatment.

This study was conducted with 1,639 individual students in 123 sixth and seventh grade classrooms from three suburban school districts in the Cleveland, Ohio, area (seven schools from across these districts were included in the study). These three school districts were selected because they had large numbers of sixth and seventh grade classes available for assignment to this study and a large student body to test the effectiveness of the study interventions. Given that the main question was whether treatment was more effective than no treatment, the team randomly assigned about half of the classes (54%) to the control condition and the other half to receive an intervention (either the interaction-based or law and justice curriculum). The research team randomly assigned the 123 study classrooms to one of three conditions:

- *Treatment 1*, an *interaction-based* curriculum that addressed GV/H by focusing on setting and communicating boundaries in relationships, the formation of respectful and mutual relationships/friendships, and the role of the bystander as intervener. Twenty-three percent of the 123 classrooms (n=28) were assigned to this intervention, which was, in most classes, conducted over a 5-week period (once per week).

- *Treatment 2*, a *law and justice* curriculum that addressed GV/H by focusing on laws, definitions, information, and data about penalties for sexual assault and sexual harassment, as well as results from research about the consequences for perpetrators of gender violence. Twenty-three percent of the 123 classrooms (n=29) were assigned to this intervention which was, in most classes, conducted over a 5-week period (once per week).

- *Control group* that went through the normal class schedule and did not receive any of the elements of treatment 1 or treatment 2. Fifty-four percent of the 123 classrooms (n=66) were assigned to this condition. The randomly assigned classes that received the control group had their regular teachers instruct their normal class, except for the days when the research team conducted the surveys. The control group completed all three waves of data collection during one of their normally scheduled periods.

Our findings from our Cleveland experiment indicated that compared to the control group, students in the law and justice treatment program had significantly improved self-reported outcomes in awareness of their abusive behaviors, attitudes toward DV/H and personal space, and knowledge of DV/H laws and resources. Compared to the control group, students in the interaction-based treatment also had many self-reported positive outcomes, including lower

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

rates of victimization, increased awareness of their abusive behaviors, and improved attitudes toward personal space requirements. Neither program affected the self-reported experience of being a perpetrator or victim of sexual harassment, student interventions as a bystander, or behavioral intentions to reduce/avoid violence. While the intervention reduced self-reported peer violence victimization and self-reported perpetration on some of the measures in these areas, there was a conflicting finding regarding self-reported dating violence perpetration. The intervention seemed to increase self-reported dating violence perpetration for some of the measures in this area (but not self-reported dating violence victimization). Nevertheless, our team's research was important because it demonstrated, through an experiment, that a condensed (compared to the much longer Fourth R and Safe Dates curricula) five-session school curriculum could be effective for a group as young as 6th and 7th grade students. However, it was unknown whether our intervention would display similar positive effects in other cities larger than the mostly suburban area outside of Cleveland as tested in this earlier project.

From 2008 to 2010, in New York City (NYC) middle schools, the team conducted a second experiment (also funded by DOJ) (Taylor et al., 2013a). In this second experiment (hereafter referred to as "NYC-1"), the researchers added a building-wide intervention component (Shifting Boundaries Schoolwide; SBS) to go along with the most effective components of the Cleveland classroom-based interventions to form a four-celled experiment testing the effectiveness of the classroom SB curriculum (SBC), SBS, both (SBC+SBS), or neither. Our team employed a multi-level, experimental design. Our team randomly assigned the SBS and SBC interventions through a stratification process with 30 public middle schools. Schools were classified by two stratifying criteria (school size and borough), ensuring that the comparison groups started out with some identical characteristics and adequate numbers of schools in each of the study cells. The schools were assigned to one of the four cells: (1) receive the SBS and SBC interventions, (2) receive SBS only, (3) receive SBC only or (4) control group (in which schools/students experienced their normal class schedule, without receiving any elements of our classroom or building-level interventions). Within each of the four cells, a random sample of classrooms was selected for study participation to complete all three waves of the survey. With the exception of three schools (which had three instead of four classrooms in the study), each school included two sixth and two seventh grade classrooms in each building, yielding a total of 117 classrooms (n=58 classes in sixth grade plus 59 classes in seventh grade) and 2,655 students. An average of 354 students participated in our study from each of the 30 school buildings. SBC was delivered through a six-session curriculum that emphasized the construction and expression of boundaries in relationships, consequences for perpetrators of DV/H, state laws and penalties for DV/H, and respectful relationships. SBS included the development and use of temporary school-based restraining orders ("respecting boundaries agreement"), higher levels of faculty and security presence in areas identified by students and school personnel as unsafe "hot spots," and the use of posters to increase awareness and reporting of DV/H to school personnel.

The main findings from this study were that the building-only and the both (classroom lessons with building-wide) interventions were effective at reducing dating and peer violence

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

victimization and perpetration (Taylor et al., 2013a). Compared to the control group which received no interventions we found the following:

- The combination of the classroom and building interventions and the building intervention alone statistically reduced sexual harassment (victimization and perpetration) by 26-34% six months post follow-up.
- The building intervention statistically reduced victimization and perpetration of physical and sexual dating violence by about 50% up to six months after the intervention.
- The combination of the classroom and building interventions and the building intervention alone led to 32-47% statistically lower peer sexual violence victimization and perpetration up to six months after the intervention.
- While the preponderance of results indicates that the interventions were effective in reducing violent incidents, a few anomalous results (e.g., reported declines in total peer violence frequency which were contradicted by higher prevalence estimates) did emerge. However, after careful analysis these anomalous results were deemed to be most likely spurious.

Overall, the SBS intervention and the SBC+SBS intervention were effective at reducing DV/H. Our findings on reducing sexual violence are noteworthy. SB was one of only a couple of interventions identified in a recent systematic review by the Centers for Disease Control and Prevention (DeGue et al., 2014) as having rigorous data to support a claim that it is successful at reducing sexual violence. Also, the success of the "building only" intervention is particularly important because it can be implemented with very few extra costs to schools. However, classroom sessions alone were not effective. Finally, our focus groups confirmed that the interventions were implemented as planned and straightforward to implement, teachers liked and were supportive of the interventions, and the positive survey results related to the interventions effectiveness were confirmed.

The current study

While some important findings were produced in the NYC-1 results, a couple of key questions remained. First, whether the SB program was of sufficient dosage to produce sustained effects post intervention beyond the six months demonstrated in NYC-1. Second, if schools could implement the SB program in just one grade to conserve resources but still achieve DV/H reduction effects. While the Safe Dates evaluation (Foshee, Bauman, et al., 2004) had assessed the question of dosage (and found no additional TDV reductions associated with a booster session), we did not find another study in the literature that explored the effects of saturation of an intervention across the middle school grades versus one grade receiving a DV/H intervention.

In 2011, the same team started this third experiment (referred to hereafter as the "NYC-2" study) funded by NIJ/DOJ. This new study extends the earlier work in Cleveland and NYC-1 by: Expanding the study to include 8th grade as well as 6th and 7th grade students, including the use of a few lessons from Safe Dates for 8th graders (Foshee et al., 1998) to test revised grade-specific DV/H interventions for middle schools (our earlier work used non-grade differentiated classroom lessons), and having a longer follow-up period of 12 months instead of the 6 months

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

follow-up of the NYC-1 research. NYC-2 included both the SB classroom curriculum (SBC) and SB school (building-level) (SBS) components in all of the treatment cells. Overall, this study helps address several gaps in the literature regarding how often to intervene, and how broadly to intervene across grade levels in order to produce a safe environment.

Our first objective was to assess whether and how much of a difference it makes when all three grades in a middle school receive a set of DV/H classroom and building-based interventions compared to when only two of the three grades receive it or only one grade receives it. To address this question, we built on our earlier research to assess the impact of *saturating* a middle school environment with information and behavioral strategies tailored to each grade level—a departure from our earlier work that looked at only the 6th and the 7th grades—acknowledging the integrated social environment of a middle school that may contribute to ongoing dating violence. Our second objective was to examine the impact of multiple *doses* of grade-differentiated curricula and whether additional dosages over two years of SB leads to greater reductions in DV/H than single dosages in one year.

Methods

The purpose of this randomized multi-level experiment (data collected from December 2011 to September 2014) was to provide high-quality scientific evidence concerning the comparative effectiveness of targeting a young, universal primary prevention audience with a multi-level intervention (the SBC classroom curricula and the SBS building-wide environmental interventions) of different saturation and dosage levels. The classroom intervention was delivered through a multi-session curriculum that emphasized the consequences for perpetrators of DV/H, state laws and penalties for DV/H, and respectful relationships. The school (building-level) based intervention included the development and use of temporary school-based restraining orders, higher levels of faculty and security presence in areas identified by students and school personnel as unsafe "hot spots," and the use of posters to increase awareness and reporting of DV/H to school personnel. The study was designed to yield data that could help increase the capacity of schools to prevent DV/H.

To follow we provide a description of our intervention, the rationale for our research site location in New York City, our use of an experimental design, and how we overcome a series of challenges in implementing our experiment.

Description of Interventions

The intervention we tested had two main components. First, we had an SB classroom curriculum (SBC), covering the consequences for perpetrators of DV/H, laws and penalties for DV/H, and respectful relationships. Second, we had an SB school (building-level) (SBS) component which included the use of school-based restraining orders, higher levels of faculty and security presence in areas identified through student mapping of safe/unsafe "hot spots," and the use of posters to increase awareness and reporting of DV/H.

The SB program uses the concept of boundaries as the fundamental building blocks through which students begin to understand concepts such as sexual harassment, dating violence,

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

sexual assault, and consent. These concepts in fact have laws (and laws are types of boundaries). The articulation of boundaries allows students to learn that they have rights and that they are entitled to assert those rights, sometimes with the force of law behind them.

Our team developed the interventions with significant input from the New York City Department of Education (NYC DOE) central office and regional office personnel. First, we held meetings with a small group of prevention content experts from the NYC DOE to gain their feedback on the lessons and to look for approaches to differentiate the lessons by grade and insert relevant local terms and expressions that are used in the NYC area. As we learned from the Taylor et al. (Taylor et al., 2008) study, input from local school personnel proved to be essential prior to the piloting testing and at the conclusion of the pilot testing. Incorporating school personnel feedback at all decision points helped shape our interventions in a way that best suited the students in NYC. The lessons were implemented by school counselors. The program component of our project team trained the school personnel at a day-long in-person training in the lessons and building-level interventions. Project staff also answered follow-up questions from the trainees over the phone and conducted refresher trainings as needed.

Classroom-based intervention (SBC). At the outset of this project (NYC-2) we worked on some refinements to the lessons and differentiating the lessons by grade. We started with the SBC lessons from our prior NIJ-funded study in New York City in 2010 (Taylor et al., 2013a) that were based on an earlier version of the SB curricula tested in our Cleveland area (2005-2007) experiment (Taylor et al., 2008). Our classroom lessons were based on synthesizing a set of lessons that proved to be most successful in the Cleveland area from the Interaction–based treatment and our Law & Justice Treatment (LJT). Our merged multi-session curriculum emphasized the development and articulation of boundaries, consequences for perpetrators of DV/H, state and federal laws related to DV/H, and respectful relationships.

For the current NYC-2 project, our SBC classroom lessons varied in the level of content and length by grade: sixth grade has four lessons, seventh grade has six lessons, and the eighth grade has seven lessons. The additional lessons for the 7th and 8th grade students include more advanced material and content for this older group adjusted to be developmentally appropriate. A key SBC goal was for the interventionists to be able to complete the teaching of the lessons in a relatively short amount of time. While there are comparative trade-offs and potential advantages of a longer more comprehensive program, we wanted to assess a curriculum that reflected the realities of limited class time for this type of effort, a sentiment that our team heard from educators across the nation (including educators in New York City).

Drawing from the previous NYC-1 SB curriculum, 7th grade students were provided lessons that emphasize the consequences for perpetrators of DV/H, state and federal laws for DV/H and sexual harassment, the setting and communicating of one's boundaries in interpersonal relationships, and the role of bystanders as interveners. The 7th grade lessons were based on our earlier experimental research in Cleveland on the effectiveness of activities exploring the concepts of laws/boundaries, plotting the shifting nature of personal space, considering laws as they apply by gender in "Big deal/No big deal," and a final activity on sexual harassment

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

through the "Says Who" quiz (see Appendix 1 for our classroom curricula we used for this second study in NYC).

As we have learned from our past work, we needed to tread more gently with the 6th grade students as we opened discussion about setting and articulating boundaries. We still addressed the topics covered with 7th graders, but with 6th graders in a more incremental, basic approach, with extra attention to introductory material.

Our 6th grade lessons concentrated on the definitions and applications of "personal space" and "boundaries"— notions that are synonymous with laws—distinguishing permissible behaviors from those that are not. From the obvious wall that serves as a boundary of a room, to a border that delineates one state or nation from another, to a more abstract use of boundaries employed in rule- and law-making, students had various opportunities to apply these concepts in activities themed around precursors to DV/H. The concept of boundaries is foundational to the entire set of SBC lessons, and the first lessons offered open-ended questions which gave students opportunities to reveal their thinking at the beginning of the intervention.

It has been our experience that students in this age group swing between concrete operational thinking to hypothetical and abstract thinking. Accordingly, we used lessons that employ both concrete, applied materials (such as "mapping safe and unsafe spaces" and "measuring personal space"), as well as activities that offer more abstract thinking. For the 7th and 8th graders, the curriculum also included a fact-based component based on the idea that increased knowledge about facts and consequences of one's behaviors are appropriate and effective primary prevention tools. These lessons present facts and statistics about sexual harassment, sexual assault and dating violence, including legal definitions of sexual assault, definitions of the different types of abuse, how to help a friend, and resources for help. Students move from discussions of measuring personal space to behaviors that are against school rules, to behaviors that are against the law.

Focus groups in our NYC-1 study with 6th and 7th grade students in mid-April 2010 reaffirmed an overall sense of urgency to include 8th grade students in future versions of our research (Taylor, Stein, Woods, & Mumford, 2011). Without exception, all of the students included in our three focus groups (N = 30), boys and girls alike, indicated that 8th grade students were the ones most likely to harass them (as younger students) and were also the ones who lacked the new frame of reference that involvement in our lessons had provided to the younger students (Taylor et al., 2011). Their ringing endorsement of our lessons and interventions provided us with new material (qualitative data) that informed the essential rationale for varying our treatment of the three middle school grades in the current NYC-2 study.

Eighth grade students received some similar material as the 6th and 7th grade students but also received additional lessons based on the TDV curriculum called *Safe Dates*, which was designed and experimentally tested for use with 8th grade students (Foshee et al., 1998;

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Foshee, Bauman, et al., 2004). Drawing from the SB lessons used with the younger grade students, the 8th grade lessons included additional material on finding and articulating personal space, establishing boundaries in relationships, mapping safe and unsafe areas of the school, and from Safe Dates - recognizing caring relationships, identifying harmful behaviors in dating relationships, the consequences of harmful behaviors in dating relationships, and helping friends (Foshee & Langwick, 2010). The reader is referred to the Hazelden publication *Safe Dates: An Adolescent Dating Abuse Prevention Curriculum Manual (2nd Edition)* for any copyrighted material used for the 8th grade Shifting Boundaries program (Foshee & Langwick, 2010).

School (building-level) based intervention (SBS). As recommended at an NIJ/National Institutes of Health (NIH) teen dating violence expert meeting (Dec. 4-5, 2007), multi-level interventions were included in our research. Specifically, we included a building-level intervention, as we did in our NYC-1 experiment (Taylor et al., 2013a), that included the following five components. The **1st** component was revised school protocols for identifying and responding to DV/H. The **2nd** component was the introduction of temporary school-based restraining orders (SBRO) (see Appendix 1 for a sample of our SBRO called a *Respecting Boundaries Agreement* or RBA). The **3rd** component was the placement of teen dating violence prevention posters in multiple locations around the school, including hotspots, with contact names for school counselors to increase awareness and reporting of violence to change environment of school to no tolerance for violence.

In addition, building on research by Astor, Meyer and others (Astor, Meyer, & Pitner, 2001; Astor, Meyer, & Behre, 1999), our intervention includes a **4th** component to help schools work with students to identify any unsafe areas of schools through hotspot mapping. School counselors or designated teachers worked with representative groups of students to identify "hotspots" where students feel safe and unsafe. Students used a blue print or other rendering of the school grounds and colored in the places that they felt unsafe (red), safe (green), and unsure of safety (yellow) with colored pencils. The map results are then aggregated and used by the school's safety and supervisory personnel to adjust security and supervisory operations to the school "hot spots" and improve the school environment. To promote greater comparability across our interventions, we applied the same basic "dosage" for the building intervention as we applied for the classroom lessons. That is, the building interventions were conducted for the same number of weeks as the classroom-based intervention (about six to ten weeks). A **5th** and final component was the adjustment of school security and supervisory personnel based on the location of the "hot spots" and awareness raising among the school educators about the "hot spots."

Fidelity of intervention implementation

We had project staff (mostly research assistants and one of the principal investigators) call the lead interventionist in each school on a regular basis (weekly during field implementation) to make sure they were able to implement the program (e.g., classroom lessons or redistribution of security staff as part of the hotspots mapping intervention) and if they had any questions on aspects of the intervention they needed clarified. While we asked the

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

interventionists to also complete project logs, we found the phone calls to more revealing. Given their busy schedules in the schools, some of the interventionists tended to complete their logs without much detail. However, we were often able to capture these details on short phone calls. The principal investigators also conducted site visits to directly observe the interventions taking place. These visits provided useful qualitative data on the nature of implementation and identified any concerns the interventionists might have had in carrying out the SB program. In general, we found that for the schools that stayed in the program there were few concerns with implementation. As documented late in this report, the larger issue was all of the schools that completed dropped out of the study or even failed to ever get started with implementation. Given the lack of variation in the schools that did participate in the project (all implemented the full SB program), there was no need to add control variables into our statistical models for varying levels of implementation.

Research Site Location

Partnering with the New York City (NYC) Department of Education (DOE) offered a rare opportunity to conduct our experiment with the largest school district in the U.S. NYC not only has the requisite number of middle school buildings called for in our design, but it also comprises one of the most ethnically, linguistically, and racially diverse populations in the U.S. For example, during the timeframe of our study, the racial makeup of students across the city was 40% Hispanic, 30% African-American, 15% Asian, 14% white and about 1% other. Serving over 1.1 million students, the system has over 1,800 schools, employs 80,000 teachers, and operates on an annual budget of $25 billion. Because of its immense size, the NYC Public School System is one of the most influential in the U.S. New experiments in teacher training and classroom pedagogy often originate in New York and then spread to the rest of the country.

Experimental Design

Our design responds to the call of policymakers to conduct rigorous research. Among the flaws found in the DV/H prevention program literature are some earlier studies with non-comparable comparison groups (for example, (Hilton, Harris, Rice, Krans, & Lavigne, 1998; Jaffe et al., 1992). The best of these studies have attempted to draw comparison groups in ways that maximize the likelihood that they will be similar to the treatment group. Our team (Taylor et al., 2013a; Taylor, 2010) and a few other researchers (Foshee et al., 1998; Peskin et al., 2014; Wolfe et al., 2009) have conducted randomized experiments with middle school students on DV/H prevention, and there have been a few quasi-experiments (QEs) with matched control groups (Jones, 1991; Krajewski, Rybarik, Dosch, & Gilmore, 1996; Weisz & Black, 2001). The problem with the QEs is that although measured differences can be statistically controlled, the many unmeasured variables related to the outcome variable (e.g., motivation to change) cannot be controlled. Randomized control trials (RCTs) are typically considered the best method for eliminating threats to internal validity in evaluating social policies and programs (Berk, Boruch, Chambers, Rossi, & Witte, 1985; Boruch, McSweeny, & Soderstrom, 1978; Campbell, 1969; Campbell & Stanley, 1963; Dennis & Boruch, 1989; Riecken et al., 1974). When RCT results are contrasted with results from other major designs and statistical alternatives, different effect sizes are found (Fraker & Maynard, 1987; Lalonde, 1986). Also, the variation in results across the QEs is greater than across the RCTs (Lipsey & Wilson, 1993).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Our study employed a multi-level, experimental, longitudinal design. The study included 23 middle schools, four treatment groups and three waves of student self-report surveys (baseline, 6-month follow-up and 12 month follow-up). We examined schools that we provided varying levels of dosages of SB (inclusive of SBC and SBS interventions):

- **Group 1**: schools assigned here received SB in one school year for 6th graders only,
- **Group 2**: schools received SB in one school year for their 6th and 7th grades,
- **Group 3**: schools received SB in one school year for their 6th, 7th and 8th grades,
- **Group 4**: schools received SB over two school years first in 6th grade and the same group receives it in 7th grade the following school year.

The unit of assignment and unit of analysis were schools. With this type of design (students nested within schools), we added a statistical correction in our models to provide for robust clustered standard errors (see "Analysis Section").

Schools were assigned to conditions according to SAS computer-generated random numbers (Shadish, Cook, & Campbell, 2002). Logistically, it would not have been possible to assign individual students to the four cells or groups of our experiment, as described immediately above, for that would require taking them out of their regular class schedules. Also, randomly assigning at the classroom level for an experiment that includes a building-wide condition opens up the possibility of contamination concerns. The strengths of our design were that there was very little possibility of contamination across the four cells. That is, the classrooms and students from the four cells were in different buildings with little opportunity for contact. Additionally, the management of this project was simplified, as each building was dedicated to one of the four assigned cells. Despite the simplicity of our design, procedures were put in place to monitor the integrity of the school and classroom SB implementation process (and monitor for expectancy, novelty, disruption, and local history events) and to measure and control for any contamination. Also, later in the analysis section, we present data demonstrating that the experiment achieved its basic purpose of creating comparable conditions to assess outcome differences in our different groups. That is, while we found a few small differences across the treatment groups prior to the experiment (during the baseline period), the four study groups/conditions were very similar on the vast majority of our measures, leaving the only major differences across the groups their assigned intervention condition.

Challenges in Implementing our Experiment:

Our own experience and that of others have shown that conducting experiments in field settings is a challenging undertaking (Davis & Taylor, 1995; Davis & Taylor, 1997; Sherman, 1992; Taylor et al., 2008). The biggest problem we had was getting schools to agree to participate in our experiment. We contacted over 75 schools to participate in our study and 35 agreed to be a part of the experiment. However, among the original 35 schools that agreed to participate and were randomized to one of the study conditions we had 12 schools dropout of the study: One of the dropped schools was from Group 1 (6th grade only), five from Group 2 (6th and 7th grade), two from Group 3 (6th, 7th and 8th grades) and four from Group 4 (6th grade

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

longitudinal). However, on key aggregate school-level statistics — size of student population, student-to-teacher ratio, percent of students receiving free or reduced priced lunch, race/ethnicity, number of suspensions and percent of student body meeting/exceeding math or reading proficiency standards — we found no statistically significant differences between the schools that participated and the schools that ended up dropping out.

School participation in the project was a challenge for several reasons. One challenge arose from a budget cut by the State of New York and the NYC School Department of Education (known as NYC-DOE). Our project started in January 2011 and our original plan called for our team to implement the interventions with the NYC Substance Abuse Prevention Intervention Specialists (SAPIS) based in each school in the spring of 2011. Coincident with the launch of the NYC-2 experiment, we were informed that the majority of the SAPIS were to be laid off by the NYC-DOE. Without a centralized team of interventionists (such as the SAPIS program) to support the project, we instead needed to focus on assembling interventionist teams in each participating school (e.g., in some cases we had school teachers and in other cases school counselors).

Also, the NYC Mayor and Chancellor of the NYC Schools had instituted a new decentralized approach to the management of the NYC schools to foster greater empowerment of school principals. The effect of this policy change was that the Central office staff of NYC DOE who had helped us in our earlier (2008-2010) NYC-1 study would not be allowed to assist our team in identifying school leaders who might be interested in joining our research project.

The NYC DOE central office staff informed our team that our only option was to approach each school principal individually and ask each of them to select school staff to implement our interventions. This turned out to be a labor intensive task and very time-consuming. Principals are very busy running their schools and have no time to respond to requests from researchers. The key problems with the approach of individual school recruitment is that (1) there are no ways to guarantee that our outreach efforts to principals will result in a reply from them — neither hard copy mailed invitations, email invitations, phone calls or faxing were effective, (2) individual recruitment was not as efficient as working with a centralized operation such as the SAPIS program (which had been managed by the central office of the NYC DOE). These two limitations meant that we had to contact individual principals, mostly one-by-one. On top of all this, there was the Hurricane Sandy disaster which occurred over part of our field period that precluded participation by some schools.

In sum, we faced five significant challenges to school recruitment and consequently the project resources: <u>One</u>, the lay-offs of the SAPIS as resource teachers. <u>Two</u>, the project timeline was pushed back by about two years (leading to higher costs). <u>Three</u>, the project PIs (Taylor and Stein) had to talk with each school individually and recruit them one at a time to the study (as opposed to the earlier approach the team used in NYC-1, through which the NYC DOE Central Office pre-identified all the participating schools in a matter of weeks and made sure they attended the intervention trainings within two months). This change led to higher costs for the time of the principal investigators than originally planned (this task could not be handled

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

by junior staff, as the school principals wanted to talk only to the project leadership). Four, a number of schools expressed interest in our project and attended our training but later decided not to implement the project (this become a drain on our project resources). That is, without a meaningful screening system for identifying interested schools, a wider net was used for recruiting schools. Unfortunately, this sometimes led us to working with schools that had only a peripheral interest in the project (a fact which was not revealed until after we trained their school personnel). And five, Hurricane Sandy prevented a number of schools from continuing their participation.

Due to our recruitment and retention problems, we had to make a number of changes in our planned design. First, we had to drop the no treatment control group; with a smaller group of schools willing to participate in the study, we had to focus on the main research question of the comparative effectiveness of different levels of SB treatment, as the NYC-1 experiment had already addressed the question of treatment efficacy (treatment versus no-treatment). Also, with the lower levels of participation, we had to modify the balanced design with the same number of schools in each study condition. That is, after we experienced a large drop off of schools, we had to randomize new schools into the study as they signed on to the project, leading to the possibility (and later reality) that we would not have a balanced number of schools in each study condition. As see in Table 1, for our six-month follow-up analyses, we had 3 schools and 133 students in Group 1 (6th grade only), 3 schools and 271 students in Group 2 (6th and 7th grades), 9 schools and 862 students in Group 3 (6th, 7th and 8th grades) and 8 schools and 498 students in Group 4 (two years of treatment). Despite this variation in sample sizes across our four comparison groups, our power analyses revealed that we still had adequate statistical power (see power analysis section below) to detect differences across our comparison groups for our six month analyses. Also as seen in Table 1, for our 12-month follow-up analyses, we had 3 schools and 128 students in Group 1 (6th grade only), 2 schools and 227 students in Group 2 (6th and 7th grades), 4 schools and 225 students in Group 3 (6th, 7th and 8th grades) and 4 schools and 234 students in Group 4 (two years of treatment). With the reduced number of schools for our 12-month follow-up, our statistical power was reduced (see power analysis section) to detect differences across our comparison groups (i.e., we could find medium to large sized effects) for our 12-month analyses.

Table 1: Sample sizes for six and twelve month follow-up analyses

Group	Grades receiving SB	Number of schools receiving SB		Number of students receiving SB	
		6 Month	12 Month	6 Month	12 Month
1	6th grade only	3	3	133	128
2	6th and 7th grades	3	2	271	227
3	6th 7th and 8th grades	9	4	862	225
4	Two years of SB (6th then 7th)	8	4	498	234

Next, the original plan called for Group 4 schools to receive three years of treatment. However, too few Group 4 schools were willing to continue participation beyond two years of treatment due to competing academic demands within the schools. Similarly, we had hoped to

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

conduct a 24-month follow-up survey, but that plan had to be abandoned given the lack of willingness of schools to continue participation. The schools agreed to one-year of follow-up surveys. Nevertheless, the one-year follow-up still represented a longer follow-up period than the earlier research on SB that only had a six-month follow-up period (Taylor et al., 2008; Taylor et al., 2013a; Taylor, 2010).

In addition to the school recruitment and retention problems, our team had to tend to the usual concerns in field expeirments for contamination. The research team controlled the random assignment process and set up procedures to safeguard against non-research staff manipulating the random assignment process. Problems have been found in implementing randomization when a variety of human factors are not addressed (Berk, Smyth, & Sherman, 1988; Boruch & Wolhke, 1985; Conner, 1977). To address these concerns we piloted the random assignment procedures to test the feasibility and acceptability of the process in the semester prior to implementation of the study. Next, we analyzed the randomization algorithm and verified that the assignment was, in fact, random and that there was an absence of non-random strings (Boruch, 1997). We also explained the nature, rationale, and purpose of the randomization process to the NYC school staff to seek their input on implementation issues.

Some contamination problems could be due to the student participants (e.g., diffusion or imitation could occur if the control group learns about the treatment). There was no evidence that this occurred in our study. Our study school sites were spread out across the five boroughs of New York City and based on discussions we had with the site interventionists, there was little to no communication across the sites, and no evidence of diffusion of the intervention to the comparison sites. Other problems could be due to the interventionists (e.g., "Hawthorne effects," “compensatory equalization”). We took some additional steps to avoid contamination. The interventionists were kept blind to the study design and were not aware of the fact that different schools were delivering SB under different conditions. Program materials were carefully controlled by our research team, with strict prohibitions against the intervention sites sharing program materials. Therefore, even if a site wanted to implement a different condition than they were assigned, they would not have the materials to modify their assigned SB implementation plan. Further, we monitored the delivery of the intervention to confirm the successful independence of the experimental conditions.

Another potential problem is uncontrollable environmental changes (e.g., staff turnover). In general, because the buildings selected into the study were located in the same school district and in the same state and city, we anticipated that environmental changes would be experienced similarly across participating building sites.

Another challenge in conducting our study was collecting survey data from all the students in our sample. As discussed earlier, we collected three waves of surveys with the students. The first survey provided baseline measures for each of the treatment groups, the second survey measured short-term changes from baseline (6-months follow-up), and the third helped our team assess if the changes persisted over a twelve month follow-up period. We created a linked longitudinal analytic file that contained contemporaneous measures for each respondent

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

at each of these points in time. The advantages of a longitudinal survey include: reduction of sampling variability in estimates of change, measures of gross change for each sample unit, and collection of data in a time sequence that clarifies the direction as well as the magnitude of change among variables.

Nonresponse in a longitudinal survey creates analytical complexities. The effect of nonresponse is most pronounced when it is correlated with the objectives of the survey and may create serious biases in the analysis. We attempted to keep nonresponse to a minimum by providing flexible scheduling, and using a passive consent system. Despite our best efforts, there was some unavoidable nonresponse. We started with a sample of 3,367 participants who completed baseline surveys across 35 participating schools or 93% of the eligible students from the participating classes in the study from each grade (n= 3,620). We were able to retain 52.4% of the 3,367 baseline participants to complete (n= 1,764) a 6-month follow-up survey and 48.1% of those participants (n= 848) were in dating relationships over the 6-month follow-up period. Our sample dropped for the 12-month follow-up period (n= 917 or 27.2% participation from the baseline with complete data), mostly due to 10 schools not continuing through to the 12-month follow-up period (only 13 of 23 schools were in the 12 month follow-up period). Of these 13 schools that participated in the baseline survey, we had 1,032 participants complete surveys; 89% of these 1,032 participants (n= 917) completed a one-year follow-up survey (115 dropped out). The 10 schools that dropped out of the study between the baseline and one-year follow-up accounted for the remaining 2,220 of the 2,335 dropouts. Also, given that consent issues were addressed in the baseline survey, the only reason for non-response in these last waves of the survey was that of students not being available to take the survey (e.g., student moved out of school, student was absent on the day the survey and makeup survey date) or student refusal to take the survey (about 3% in waves 2 and 3).

To follow we provide a description of our methods used to administer the student survey, a review of our survey measures, and how we addressed a number of data analytic issues.

Student Survey Administration

Pencil-and-paper surveys were designed for students to complete, and were administered by school personnel (not involved in the intervention delivery) who were trained by a member of our research team in proper administration processes. The training consisted of a review of the study goals, objectives, activities, and background history to the project; details on the instruments and required information contained on each form; and legal aspects and procedures to protect human subjects. The survey administrators provided an orientation to students on the purpose of the survey and instructions on completing it. The procedures did not reveal the assignment process to the research staff administering the survey or the students completing the survey. Surveys were distributed at three different times: immediately before the assignment to one of the study conditions (a baseline survey), about 6-months after the treatment was completed, and about one-year after their assignment to one of the study conditions.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Surveys included a pre-numbered unique research identification number generated through a random number sequence. In addition, each survey had a removable sticker with the student's name and corresponding ID number affixed. This allowed the survey administrators to distribute surveys easily in classrooms. Students were instructed to remove the label before returning the completed surveys to the survey administrators to ensure confidentiality. This process occurred at the baseline and at both post-tests. The ID-to-name code matrix was only available to the research team and was kept in a secure location. The student surveys (see Appendix 2) were designed for optical scanning, and prior to the surveys being scanned into a database, they were reviewed for completeness, inadvertent missing data, and removal of all stray marks from the scan sheets. Scan operators conducted random samples of a portion of the scanned surveys (10% sample) to determine accuracy with raw data from the physical scan sheet.

Passive parental consent and child assent forms were addressed prior to the administration of the survey. Consent included permission for the students to complete a baseline and all of the subsequent follow-up surveys. Students were asked to return parent/guardian decline forms to the school as soon as possible (parents/guardians were told that nothing had to be done if they chose to have their child participate in the survey). During regular school hours identified in consultation with each school, consented students were asked to complete the survey in a classroom during one classroom period. The surveys took about 40 minutes to complete for students who assented on the day of the survey administration.

Measures

Sexual harassment victimization and perpetration. Sexual harassment is defined as unwanted sexual behavior portrayed as comments, images, or gestures that are sexual in nature. The term is inclusive of comments or behavior regarding someone's gender identity in terms of sexual preference or physical development.(Hill & Kearl, 2011) The surveys measured prevalence of the experience of being a victim and/or perpetrator of sexual harassment. We summed affirmative responses from the 9-item scale to create separate overall prevalence estimates of sexual harassment victimization and prevalence. These items, originally adapted from other work,(AAUW Educational Foundation, 1993, 2001; Basile et al., 2009; Fineran & Bennett, 1999) were tested and validated in a prior evaluation in a middle school population period (Taylor et al., 2008; Taylor et al., 2013a; Taylor, 2010). All of the sexual harassment measures, both victim and as a perpetrator measures, showed good internal consistency with Cronbach's alpha above .8.

Sexual and physical violence victimization and perpetration. While sexual harassment is also inclusive of extreme behavior such as rape,(Hill & Kearl, 2011) school-based prevention research has often measured sexual violence, as well as non-sexualized physical violence, on distinct scales. Measures of peer violence, both in terms of victimization and perpetration, (Taylor, Mumford, Liu, & Stein, 2015) are based on surveys developed specifically for assessing the impact of physical and sexual violence prevention programs (Foshee et al., 1998; Schewe, 2000b; Ward, 2002). Our survey included seven prevalence (yes/no) and seven frequency questions (e.g., How many times did you do this to them in the past 6 months? Zero times? 1 to

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

3 times? 4 to 9 times? 10 or more times?). The questions covered the experience of being a victim and/or perpetrator of sexual violence and physical/non-sexual violence by/of peers. *Physical violence* items included slapping or scratching; physically twisting an arm or bending back fingers; pushing, grabbing, shoving, or kicking somewhere on the body other than in the private parts; hitting with a fist or with something hard besides a fist; and threatening with a knife or gun. Because of sensitivity concerns raised by school personnel regarding explicit measurement of sexual violence in a middle school population, we were limited to two main *sexual violence* items ("pushed, grabbed, shoved, or kicked you in your private parts" and "made you touch their private parts or touched yours when you did not want them to"), similar to other research.(Foshee et al., 1998) The violent victimization measures had Cronbach's alpha scores above .8.

As with any self-reported measure, the study's survey measures had limitations. For example, students may have had trouble remembering the timing of a victimizing event, may have deliberately under-reported certain behavior (e.g., they may have been embarrassed to admit they were victimized or ashamed to admit they attacked someone else), or may have exaggerated certain behavior (e.g., over-reported the number of times they were physically abusive with a girl). Despite these potential problems, which likely were balanced across treatment conditions, confidential self-report surveys have become an accepted modality for collecting youth violence data (Foshee, 1996; Jouriles, Mueller, Rosenfield, McDonald, & Dodson, 2012; Wolfe et al., 2001) and are preferable to interviewer administered assessments which are associated with higher levels of social desirability biases (Ebesutani, Bernstein, Martinez, Chorta, & Weisz, 2011).

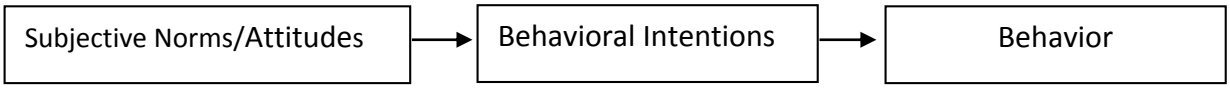
The survey also included a small number of background variables on the students, including age, gender, and ethnicity/racial background and questions on prior attendance at an educational program about sexual assault, harassment, or violence, attitudes towards DV/H, and prior history of dating.

Theoretical Framework for Interventions and Study Hypotheses

The design of SB was informed by the Theory of Reasoned Action (Ajzen & Fishbein, 1980; Fishbein, 1967). More than 40 years ago Martin Fishbein (1967) developed a versatile behavioral theory and model called the Theory of Reasoned Action (TRA). In later years, Ajzen and Fishbein (1980) refined and further specified the conditions under which behavioral change occurs. TRA emerged from prior research and theories on attitudes and later work on the relationship between attitudes and behavior. TRA addressed some of the problems with traditional attitude–behavior research, much of which found weak correlations between attitude measures and performance of behaviors (Hale, Householder, & Greene, 2003). TRA explains the main elements and inputs that result in any particular behavior. The most basic form of the TRA model[3] is the following:

[3] In 1991, Ajzen modified the model to include an interaction component called perceived behavioral control (Ajzen, 1991). This component acknowledges that there may be factors outside an individual's control that influence behavior and the intention to change or adopt new behaviors. The interaction suggests that the intention

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.



TRA proposes that your attitude towards a behavior consists of a belief that that particular behavior leads to a type of outcome and an assessment of the outcome of that behavior. If your assessment of the outcome is good you may then intend to or actually carryout such a behavior. Also a part of your attitude toward a behavior is your perceptions of what others around you believe that you should do. In the end, your attitude toward a behavior can lead to an intention to act or not act and this intention will change your likelihood of enacting a certain behavior. More specifically, TRA is based on research that demonstrates that intentions to behave are immediate predecessors to specific actions. Behavioral intentions are the proximal predictors of behavior. Based on TRA, attitudes toward and perceived norms about the desired behavior facilitate the intention to change, modify, or adopt a particular behavior.

A body of TRA-based research has emerged that suggests that people will usually act in accordance with their intentions and perceptions of control over their behavior. Over the last 30 years, TRA has helped to explain and predict behavior and has been used in a variety of contexts to better understand, for example, adolescents' intention to have sex (Fores, Tschann, & Marin, 2002), youth alcohol use (O'Callaghan, Chant, Callan, & Baglioni, 1997), smoking (McGahee, Kemp, & Tingen, 2000), drug use (Budd, Bleiker, & Spencer, 1983; Conner & Sherlock, 1998) and safer sex behaviors (De Vroome, Stroebe, Sandfort, De Wit, & Van Griensven, 2000).

The primary purpose of the TRA is to express the key factors associated with behavior change and to attempt to explain a person's behavior. While that enterprise is fraught with difficulties and is probabilistic in nature, the model served a valuable function of orienting the developer of *Shifting Boundaries* (Dr. Nan Stein) to consider the environmental context that surrounds and influences intentions and behavior. We did not set out in this project to do a formal test of TRA, but rather used it in this project to provide a framework for the development of *Shifting Boundaries*. The interventions in turn were designed to address elements of the Theory of Reasoned Action (increased knowledge is designed to change attitudes which in turn affects behavioral intentions leading to behavioral change).

---

to behave (motivation) and the ability to perform (behavioral control) combine as a meaningful predictor of change. Ajzen (1991) called the modified TRA the Theory of Planned Behavior (TPB). Ajzen also suggested that access to resources influence an individual's perceived control or power to change. Resources may include such things as support—emotional, financial as well as such daily needs as transportation to and from other resource centers. Both the TRA and TPB modification hold that the best determinant of behavior change is a person's intention to perform or not perform the behavior. The intention is influenced by multiple factors: subjective norms, attitudes towards the behavior, perceived control to engage in the behavior, and resources supportive of the desired behavior.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Data Analytic Issues

In this section, we discuss three analytic issues related to: (1) whether we had enough cases in our study to adequately detect statistical differences between the comparison groups (statistical power), (2) whether the experiment achieved its basic purpose of creating comparable treatment groups (pre-treatment study arm comparison), and (3) whether attrition in our study created any pattern of bias that would interfere with our ability to draw inferences from our study (Attrition analyses).

Following our review of these three data analytic issues we cover descriptive statistics on our sample and our main analyses (i.e., our outcome models).

Statistical power: Statistical power provides an estimate of how often one could detect a statistical relationship that in fact existed (Cohen, 1988; Weisburd, Petrosino, & Mason, 1991). We adjusted for the nesting of our multiple levels of experimental data in our power calculations. For our power calculations, we assume an alpha of 0.05, a two-tailed test, and a power level of 80%. Table 2 provides our power calculation and the minimum effect sizes we can detect given our sample sizes for our four main samples: (1) Time 1 to Time 2 analyses for our full sample of cases to explore our peer violence and sexual harassment outcomes, (2) Time 1 to Time 2 analyses for our dating sample to explore our dating violence outcomes, (3) Time 1 to Time 3 analyses for our full sample of cases to explore our peer violence and sexual harassment outcomes, (4) Time 1 to Time 3 analyses for our dating sample to explore our dating violence outcomes. Our sample size is largest for the Time 1 to Time 2 analyses for the full sample and drops when we use just our dating sample. Our smallest sample is for our Time 1 to Time 3 analyses for the dating sample.

Our smallest effects can be detected with 80% power for our Time 1 to Time 2 analyses (n= 1,763 in 23 schools) when we compare our 6th, 7th and 8th grade group (n= 862) to our combined 6 grade/6 grade longitudinal before they receive 2nd year treatment group (n= 631). Under this scenario, we have 80% power to detect an effect size as small as Cohen's D= 0.37 (a small effect size). When we conduct these same analyses at Time 3, dropping down to 814 cases in 13 schools, we have 80% power to detect an effect size as small as Cohen's D= 0.63 (a medium effect size). We have less power when we analyze our dating sample to examine our dating violence outcomes. For our Time 1 to Time 2 analyses comparing our 6th, 7th and 8th grade group of daters to our combined 6 grade and 6 grade longitudinal group of daters before they receive the 2nd year treatment we have 80% power to detect an effect size as small as Cohen's D= 0.83 (a large effect size). For our Time 1 to Time 3 analyses comparing our combined 6th and 7th grade/6th, 7th and 8th grade group to our 6 grade only group we have 80% power to detect an effect size as small as Cohen's D= 1.94 (a large effect size).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 2: Power calculation: minimum detectable effect sizes

Sample	Comparison	Physical victimization prevalence for the comparison group	minimal OR that can be detected	Cohen's D	Interpretation
T1-T2 comparison full sample (N=1764 in 23 schools)	6-8 grades (N=862) Vs 6 grade or 6 grade longitudinal (N=631)	0.54	0.52	0.37	Small effect size
T1-T2 comparison dater sample (N=848 in 23 schools)	6-8 grades (N=450) Vs 6 grade or 6 grade longitudinal (N=241)	0.14	0.22	0.83	Large effect size
T1-T3 comparison full sample (N=814 in 13 schools)	6 and 7 grades or 6-8 grades (N=452) Vs 6 grade only (N=128)	0.53	0.32	0.63	Medium effect size
T1-T3 comparison dater sample (N=443 in 13 schools)	6 and 7 grades or 6-8 grades (N=291) Vs 6 grade only (N=40)	0.17	0.03	1.94	Large effect size

What this means is that our study, with power of 80%, will find statistically significant results even when the differences between the comparison groups are between small and medium effects for our analyses of our full sample at 6 months and 12 months respectively (e.g., examining peer violence or sexual harassment). We believe effect sizes smaller than Cohen's D of .37 or .63 are not likely to be meaningful from a policy perspective. Our sample with 23 schools was also larger than the Foshee dating violence prevention experiment with 10 middle schools in rural North Carolina (Foshee et al., 2000), the Jaycox experiment in Los Angeles with 10 high schools (Jaycox et al., 2006) and the Wolfe study in Canada with 20 high schools (Wolfe et al., 2009). The statistical power of our study to find statistically significant differences in our dating sample (e.g., to examine dating violence) is not as good and means we will only be able to detect large effect size differences across our treatment groups for our dating sample.

Pre-treatment study arm comparison: The basic purpose of a randomized experiment is to create comparable conditions to assess outcome differences across the study conditions. It is always prudent when conducting RCTs to assess if this was achieved. In the analyses that

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

follow we assess (with available data) if our experiment succeeded at this task of creating comparable study conditions. Our analyses indicated that all four groups were similar regarding the following characteristics at baseline:

- Gender (X^2= 4.558, p= .21).
- Prior experience with dating violence prevention programs (X^2= 4.7, p= .19).
- Ethnicity (X^2= 3.2, p= .07).
- Number of people dated for more than one week (X^2= 2.9, p= .09).
- Length of prior dating relationships (X^2= 4.0, p= .06).
- Any peer violence physical victimization in lifetime (X^2= 4.7, p= .19).
- Any peer violence sexual victimization in lifetime (X^2= 3.3, p= .35).
- Any peer violence sexual perpetration in lifetime (X^2= 7.7, p= .06).
- Any dating violence physical victimization in lifetime (X^2= 2.1, p= .55).
- Any dating violence sexual victimization in lifetime (X^2= 3.6, p= .30).
- Any dating violence physical perpetration in lifetime (X^2= 2.1, p= .55).
- Any dating violence sexual perpetration in lifetime (X^2= 0.8, p= .86).

Despite random assignment, some statistically significant pre-treatment differences (p< .05) across the study conditions did emerge:

- The 6th grade only group had the youngest students for its sample members (46% were 11 years old) compared to the 6th, 7th and 8th grade group which only 17.6% of its sample members at 11 years old (X^2= 449.9, p< .001). Also, the 6th, 7th and 8th grade group had the oldest students for its sample members (3% were 15 years old) compared to the other groups who all had less than 1% of its members who were 15 years old (X^2= 249.1, p< .001). However, each of the four groups had a similar number of 12 year old students among their members ranging from about 30% to 40%.
- We found differences for our 6th, 7th and 8th grade group compared to the other three study conditions for peer perpetration of physical violence. Our 6th, 7th and 8th grade group had a 49% rate of peer perpetration of physical violence compared to 40% for the 6th grade only group, 44% for the 6th and 7th grade group and 39% for the multi-year treatment group (X^2= 13.2, p< .01).
- We found differences for our 6th, 7th and 8th grade group compared to the other three study conditions for sexual harassment victimization. Our 6th, 7th and 8th grade group had a 55% rate of sexual harassment victimization compared to 42% for the 6th grade only group, 46% for the 6th and 7th grade group and 42% for the multi-year treatment group (X^2= 22.6, p< .001).
- We found differences for our 6th, 7th and 8th grade group compared to the other three study conditions for sexual harassment perpetration. Our 6th, 7th and 8th grade group had a 27% rate of sexual harassment perpetration compared to 22% for the 6th grade only group, 24% for the 6th and 7th grade group and 16% for the multi-year treatment group (X^2= 24.1, p< .001).

While we found some differences between the treatment and control conditions prior to the experiment (during the baseline period), most of these differences (while statistically

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

significant) were not very large differences. For the most part, the four study groups/conditions were similar on the majority of our measures leaving the only major differences across the groups their assigned condition. Additionally, random assignment procedures were followed closely (no "overrides"). All schools assigned to treatment received their appropriate treatment. Finally, we included the variables where there were pre-treatment differences into our outcome models as covariates to remove any potential biases these small imbalances might have presented for the interpretation of our results.

Attrition analyses: Missing data can cause problems with research by reducing power and threatening the validity of statistical inferences (Fichman & Cummings, 2003). To address missing data from partially completed questionnaires, the study team used multiple imputations in the analyses. First, we created 25 multiply imputed datasets in Stata using all of the variables included in our analyses, via multivariate imputation by chained equations (MICE) approach (also known as sequential regression multiple imputation, results pooled using Rubin's rules (Van Buuren, 2007). Multiple imputation involves filling in the missing values multiple times, creating multiple "complete" datasets (Azur, Stuart, Frangakis, & Leaf, 2011). Creating multiple imputations, as opposed to single imputations, accounts for the statistical uncertainty in the imputations and using a chained equations approach adds flexibility by allowing for the handling of variables of varying (e.g., ordinal data) types (Azur et al., 2011). Single imputation procedures, such as mean imputation, are an improvement over ignoring a missing data issue but do not account for the uncertainty in the imputations; once the imputation is completed, analyses proceed as if the imputed values were the known, true values rather than imputed, leading to the potential for incorrect conclusions (Azur et al., 2011). MICE is a particular multiple imputation technique (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001) and operates under the assumption that given the variables used in the imputation procedure, the missing data are Missing At Random (MAR), which means that the probability that a value is missing depends only on observed values and not on unobserved values (Raghunathan et al., 2001). Using the MICE procedure a series of regression models are run whereby each variable with missing data is modeled conditional upon the other variables in the data, allowing each variable to be modeled according to its distribution, with, for example, binary variables modeled using logistic regression (Raghunathan et al., 2001). We analyzed our resulting imputed data sets using Stata 12 software which supports the analysis of multiple imputed data using the command "ice."

In the two sets of analyses that follow we compare **(1)** our 6th month follow-up participating schools (n=23) to those schools that dropped out of the study and only did the baseline survey (n=12) on our key baseline survey results and **(2)** our 12th month follow-up participating schools (n=13) to those schools that did not do the 12th month survey (n=22) (either because they dropped out of the study by the 12th month mark [n=10] or only did the baseline survey [n= 12]) on our key baseline survey results.

A comparison of the baseline survey data from the 23 schools that participated in the 6th month follow-up survey to those schools that dropped out of the study and only did the

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

baseline survey (n=12) revealed that there were no differences between the two groups regarding the following characteristics at baseline:

- Gender (X^2= 0.5, p= .47).
- Ethnicity (X^2= 1.9, p= .19).
- Ever in a dating relationship (X^2= 2.3, p= .13).
- Number of people dated for more than one week (X^2= 4.1, p= .66).
- Length of prior dating relationships (X^2= 0.2, p= .68).
- Prior experience with dating violence prevention programs (X^2= 0.9, p= .34).
- Any peer violence physical victimization in lifetime (X^2= 4.0, p= .06).
- Any peer violence sexual victimization in lifetime (X^2= 1.1, p= .30).
- Any peer violence physical perpetration in lifetime (X^2= 3.6, p= .058).
- Any peer violence sexual perpetration in lifetime (X^2= 1.0, p= .31).
- Any sexual harassment perpetration in lifetime (X^2= 0.85, p= .36).

A comparison of the baseline survey data from those that participated in the 6th month follow-up survey to those schools that dropped out and only did the baseline survey revealed some differences between the two groups regarding the following characteristics at baseline:

- We found small but statistically significant differences for those who did <u>not</u> complete the six month follow-up survey compared to those who did for sexual harassment victimization. Those who did <u>not</u> complete the six month follow-up survey had a higher rate of sexual harassment victimization (54%) compared to those who did the six-month survey (49%) for sexual harassment victimization (X^2= 8.9, p< .01).
- We found small but statistically significant differences for those who did <u>not</u> complete the six month follow-up survey compared to those who did for the age of the respondent. Those who did <u>not</u> complete the six month follow-up survey were less likely to be an 11 year old (21%) and more likely to be a 12 year old (44%) compared to those who did the six-month survey (11 years old was 35% and 12 years old was 33%) (X^2= 71.2, p< .001).

A comparison of the baseline survey data from the 13 schools that participated in the 12th month follow-up survey to those schools that did not do the 12th month survey (n=22) on our key baseline survey results revealed that there were no differences between the two groups regarding the following characteristics at baseline:

- Gender (X^2= 1.9, p= .17).
- Ethnicity (X^2= 1.8, p= .17).
- Ever in a dating relationship (X^2= 0.02, p= .96).
- Number of people dated for more than one week (X^2= 11.5, p= .07).
- Length of prior dating relationships (X^2= 6.9, p= .14).
- Prior experience with dating violence prevention programs (X^2= 5.3, p= .055).
- Any peer violence physical victimization in lifetime (X^2= 0.06, p= .81).
- Any peer violence sexual victimization in lifetime (X^2= 1.1, p= .30).
- Any peer violence physical perpetration in lifetime (X^2= 0.006, p= .94).
- Any peer violence sexual perpetration in lifetime (X^2= 1.8, p= .18).
- Any sexual harassment perpetration in lifetime (X^2= 0.45, p= .50).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

A comparison of the baseline survey data from the schools that participated in the 12th month follow-up survey to those schools that did not do the 12th month survey on our key baseline survey results revealed some differences between the two groups regarding the following characteristics at baseline:

- We found small but statistically significant differences for those who did not complete the 12-month follow-up survey compared to those who did for sexual harassment victimization. Those who did not complete the 12-month follow-up survey had a higher rate of sexual harassment victimization (52%) compared to those who did the six-month survey (48%) for sexual harassment victimization (X^2= 5.1, p= .024).

Overall, we did not observe much by way of patterns in our study for the schools that continued on to complete the follow-up wave surveys and those schools that dropped out after doing only a baseline survey. For the 6-month follow-up survey data, we found small but statistically significant differences in the pre-treatment rates of sexual harassment victimization (54% to 49%) and differences by age (those who did not complete the six month follow-up survey were less likely to be an 11 year old and more likely to be a 12 year old compared to those who did the six-month survey). For the 12-month follow-up survey data, we also found small but statistically significant differences in the pre-treatment rates of sexual harassment victimization (52% to 48%). In our later outcome models we include, among other variables, age and pre-treatment exposure to sexual harassment as covariates. Therefore, whatever impact these small differences might have on our outcome models are controlled for through the use of covariates. In sum, attrition analysis showed few differences between baseline participants and follow up participants, suggesting that the study sample is fairly representative of the original sample.

Results

Descriptive Statistics on the Sample

Our first sets of analyses based on our baseline data describe the key analytic variables connected with the project aims. Our study not only provided for a rigorous comparison (the evaluation component), but also provided useful descriptive information about an understudied phenomenon (DV/H) among middle school students.

Demographics. As described in the Methods section above, a total of 3,367 participants completed a baseline survey across 35 participating schools. We were able to retain 52.4% of the 3,367 baseline participants to complete a 6-month follow-up survey (n= 1,764) and 48.1% of those participants (n= 848) were in dating relationships over the 6-month follow-up period.

Participating students ranged in age from 10 to 15. Less than 1% (0.74%) were 10 years old, 33% were 11 years old, 34% were 12 years old, 24% were 13 years old, 7% were 14 years old, and 2% were 15 years old. The sample was nearly even split between boys (50.7%) and girls (49.3%), with an even distribution of boys and girls across our age categories (X^2= 9.2, p= .10).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Partnering with the NYC Department of Education (DOE) offered a rare opportunity to conduct our experiment with the largest school district in the U.S. NYC not only had the requisite number of middle school buildings called for in our design, but it also comprises one of the most ethnically, linguistically, and racially diverse populations in the U.S. During the timeframe of our study, the racial makeup of students across the city was 40% Hispanic, 30% African-American, 15% Asian, 14% white and about 1% other. While we collected our data with Hispanic as a separate question from race, for comparability purposes with overall city data we assembled the following: In our sample, we had a fairly close ethnic breakdown to the overall city average of 26% Hispanic, 37% African American, 16% Asian, 14% white and the remainder (7%) in the "other" racial category.

Prior education in violence prevention. Over a third of the study sample (44.3%) had prior experience with a violence prevention educational program. The item queried experience with "*educational program[s] about sexual harassment, sexual assault/rape, dating violence, and/or family/domestic violence.*" However, the specific nature of that educational program and the extent to which it addressed peer or dating relationship violence was not explored.

Prior relationship experience. Nearly half of the sample (47.7%) reported at least one experience of being in a dating relationship that lasted one week or longer. The majority of those who report having dated had at least 2 partners (35% reported one prior partner, and 22% reported two partners) in their lifetime. Also, 13% of our sample report having three partners. Only 15% of our sample reported 7 or more partners in their lifetime. The relative frequency of relationships for middle school students is tempered by their short duration: only 29.6% of students who reported ever having been in a dating relationship indicated that they had been in a relationship that lasted more than six months (11.3% of our sample were in relationships of about 1 week, 22.9% greater than 1 week and less than one month, 36.2% from 1 month to 6 months), with only 11.7% in relationships greater than 6 months and less than a year and 18% of our sample in relationships greater than one year.

Prior experience of victimization. The project collected data on three main forms of victimization: dating violence (any physical and sexual violence), peer violence (any physical and sexual violence), and sexual harassment. One in five respondents (19.8%) reported having been the victim of any physical dating violence. One in ten respondents (9.5%) reported having been the victim of any sexual dating violence. Almost 60% of the sample (59%) reported having been the victim of any physical peer violence at some point in time, and 18.1% were the victim of sexual peer violence at some point in time. Also, 49% report having been sexually harassed at some point in time.

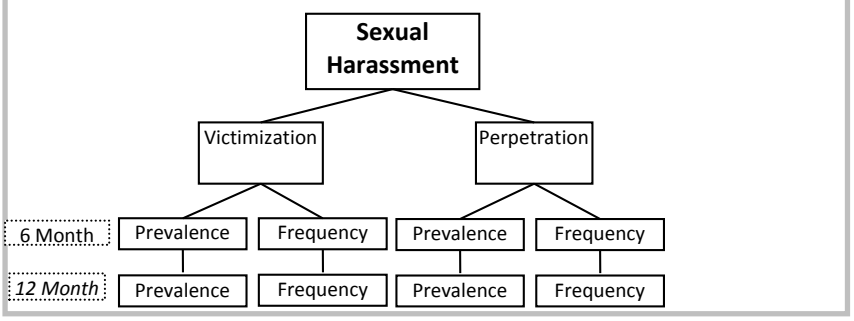
Prior experience of perpetration. The project collected data on three main forms of perpetration: dating violence (any physical and sexual violence), peer violence (any physical and sexual violence), and sexual harassment. One in five respondents (19%) reported having perpetrated any physical dating violence at some point in time. About 6% (6.4%) of the respondents reported having perpetrated sexual dating violence at some point in time. Nearly half (45%) reported having perpetrated any physical peer violence at some point in time and 8%

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

reported having perpetrated sexual peer violence at some point in time and nearly a quarter (23%) report having sexually harassed someone at some point in time.

Outcome Models

Figure 1: Models of Sexual Harassment Outcomes (8 models)



As described earlier, given our use of a clustered randomized trial, we needed a statistical technique to address the clustered nature of our data (students nested within schools). This is a concern because variables at the student-level and school level may be correlated (i.e., not independent). In the past, hierarchical data were analyzed using conventional regressions, but these techniques yield biased standard errors and sometimes spurious results (Hox, 2002). Also, analyzing only at the aggregate level will lead to a loss of information and power. As early as 1978, experimental researchers noted, ''analyses of group randomized trials that ignore clustering are an exercise in self-deception'' (Cornfield, 1978). To address this concern, we added a statistical correction in our models to provide robust clustered standard errors which adjusts for within-cluster correlation through including a robust variance estimate.

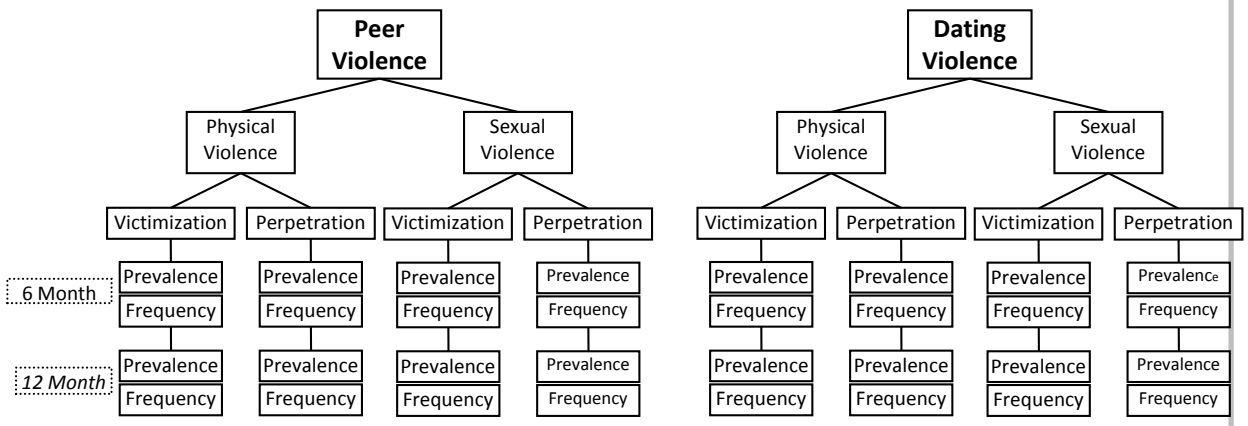
We conducted these analyses using Stata 12.0 statistical software with the vce (cluster *clustvar*) option. The robust variance estimator comes under various names in the literature, but within the Stata software it is known as the Huber/White/sandwich estimate of variance (Froot, 1989; Rogers, 1993; Williams, 2000). The names Huber and White refer to the seminal references for this estimator (Huber, 1967; White, 1980). The name "sandwich" refers to the mathematical form of the estimate, namely, that it is calculated as the product of several matrices.[4] Given our substantive interest in the individual data, and our need to only control for the clustering of the school-level data, the use of a robust variance estimator to address the

[4] The matrix formed by taking the outer product of the observation-level likelihood/ pseudo-likelihood score vectors, used as the middle of these matrices (the meat of the sandwich), and this matrix is in turn pre- and post-multiplied by a model-based variance matrix (the bread of the sandwich) (Rogers, 1993). The robust calculation is generalized by substituting the meat of the sandwich with a matrix formed by taking the outer product of the *cluster-level* scores, where within each cluster the cluster-level score is obtained by summing the observation-level scores (Rogers, 1993; Williams, 2000; Wooldridge, 2002).

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

clustered nature of our data and produce unbiased estimates was adopted (Rogers, 1993; Williams, 2000).[5]

Figure 2: Models of Violence Outcomes (32 models)



To follow is a presentation of our outcome models in the following areas for Wave 2 (6-months post treatment) and Wave 3 (about 12 months post treatment): sexual harassment (8 models as depicted in Figure 1 above), and violent behavior (16 models for peers and for dating relationships as depicted in Figure 2 [total 32 models]). Our six-month follow-up data allows us to address the question of how much of a difference it makes when only the 6th grade in a middle school receives SB compared to when the 6th and 7th grade in a middle school receive SB or when all three grades in a middle school receive SB? Our twelve-month follow-up data allows us to address the question of whether additional dosages of SB leads to greater reductions in DV/H than single dosages (6th grade longitudinal versus 6th grade only groups) and assess our first saturation-level question on the relative value of multiple grades receiving SB compared to one grade at the 12-month follow-up mark.

As discussed in the methods section earlier, we had to modify our planned design for this experiment due to low levels of participation among the New York City middle schools. While we planned to have a no-treatment control group that proved not feasible in this context, we decided to maximize our use of the schools to address our main research question of the comparative effectiveness of different levels of SB treatment. Also, we had a no treatment control group with our earlier NYC experiment conducted only a couple of years before the current study and it already addressed the question of treatment efficacy (treatment versus no-treatment).

Six-month follow-up outcome data. Our six month follow-up data allows us to address the saturation question of how much of a difference it makes when only the 6th grade in a middle school receives SB compared to higher levels of saturation when the 6th and 7th grade in a

[5] Given our need to adjust for clustered standard errors through a modeling based approach, we do not present simple means for the treatment and control groups.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

middle school receive SB or when all three grades in a middle school receive SB. For these analyses we combined our 6th grade only group (n= 3 schools) with our 6th grade longitudinal group (n= 8 schools) because at the sixth month follow-up mark the 6th grade longitudinal group had only received the 6th grade intervention and was functionally the equivalent of the 6th grade only group. The combined group had 11 schools and 631 students. Our next group is made up schools assigned to receive the 6th and 7th grade SB intervention (n= 3 schools and 271 students). The final group is made up schools assigned to receive our 6th, 7th and 8th grade SB intervention (n= 9 schools and 862 students).

In this section we present four tables. Each model in each of the four tables (Table 3-6) includes Adjusted Odds Ratios (AOR) (for the prevalence outcomes) or Regression Coefficients (for the frequency outcomes) and their 95% confidence intervals (in parentheses) with p values for each treatment arm, with the 6th grade only group serving as the reference group in each model. Each model includes covariates for gender, age, sexual harassment victimization and perpetration and a control for clustering of students within schools. We did not control for other variables because they were not statistically significant in our earlier models.

Table 3 (see below) covers the prevalence of peer violence and sexual harassment at the six-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the 6th and 7th grade group and the 6th, 7th and 8th grade group were statistically significant, suggesting that the treatment effects were equal.

Table 3: Six-month intervention effect on **prevalence** of peer violence and sexual harassment: Adjusted Odds Ratio controlling for gender, age, and wave I sexual harassment victimization and perpetration and controlling for clustering of students within schools (N=1764)*

Outcome	Treatment conditions (6 grade only as reference)	AOR (95% CI)	*p*
Peer physical victimization	6, 7 and 8 grades	0.93 (0.63,1.37)	0.70
	6 and 7 grades	0.77 (0.47,1.26)	0.28
Peer sexual victimization	6, 7 and 8 grades	0.77 (0.45,1.31))	0.31
	6 and 7 grades	0.77 (0.33,1.77)	0.52
Peer physical perpetration	6, 7 and 8 grades	1.12 (0.75,1.67)	0.55
	6 and 7 grades	1.16 (0.66,2.02)	0.60
Peer sexual perpetration	6, 7 and 8 grades	0.72 (0.40,1.30)	0.26
	6 and 7 grades	0.91 (0.27,3.07)	0.88
Sexual harassment victimization	6, 7 and 8 grades	0.95 (0.58,1.53)	0.81
	6 and 7 grades	0.77 (0.42,1.41)	0.38
Sexual harassment perpetration	6, 7 and 8 grades	0.87 (0.53,1.44)	0.58
	6 and 7 grades	0.86 (0.47,1.59)	0.61

*reference intervention group: 6 grade only

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 4 (see below) covers the frequency of peer violence and sexual harassment at the six-month post intervention mark. We had one statistically significant effect (i.e., $p < .05$) in Table 4. The 6th & 7th grade dosage was associated with less sexual harassment victimization frequency compared to the reference category of 6th grade only (Regression Coefficient = -0.18, p= 0.01). The standardized effect size for this finding was a Cohen's D of 0.20 which is equivalent to a small effect just below a medium sized effect (Cohen, 1988). We also had two borderline cases (p<.10). The 6th & 7th grade group was associated with less peer physical victimization frequency compared to the reference category of 6th grade only (Regression Coefficient = -0.12, p= 0.08). The 6th & 7th grade group was associated with less peer sexual perpetration frequency compared to the reference category of 6th grade only (Regression Coefficient = -0.06, p= 0.09). However, none of the other treatment comparisons between the reference category of the 6th grade-only group to the 6th and 7th grade group and the 6th, 7th and 8th grade group were statistically significant.

Table 4: Six-month intervention effect on **frequency** of peer violence and sexual harassment: Regression coefficients controlling for gender, age, and wave I sexual harassment victimization and perpetration and controlling for clustering of students within schools (N=1764)*
*reference intervention group: 6 grade only or 6 grade longitudinal

Outcome	Treatment conditions (6 grade only as reference)	6 months post intervention Regression Coefficient (95% CI)	*p*
Peer physical victimization	6, 7 and 8 grades	0.03 (-0.10,0.16)	0.64
	6 and 7 grades	-0.12 (-0.25,0.01)	0.08
Peer sexual victimization	6, 7 and 8 grades	-0.02 (-0.10,0.06)	0.62
	6 and 7 grades	-0.04 (-0.14,0.06)	0.37
Peer physical perpetration	6, 7 and 8 grades	0.06 (-0.05,0.17)	0.28
	6 and 7 grades	0 (-0.13,0.13)	0.97
Peer sexual perpetration	6, 7 and 8 grades	-0.07 (-0.16,0.03)	0.15
	6 and 7 grades	-0.06 (-0.14,0.01)	0.09
Sexual harassment victimization	6, 7 and 8 grades	0.02 (-0.15,0.18)	0.83
	6 and 7 grades	-0.18 (-0.33,-0.04)	0.01
Sexual harassment perpetration	6, 7 and 8 grades	0 (-0.10,0.10)	0.93
	6 and 7 grades	-0.03 (-0.12,0.06)	0.49

*reference intervention group: 6 grade only or 6 grade longitudinal

Table 5 (see below) covers the prevalence of physical dating violence and sexual dating violence at the six-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the 6th and 7th grade group and the 6th, 7th and 8th grade group were statistically significant, suggesting that the treatment effects were equal.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 5: Six-month intervention effect on **prevalence** of dating violence: Adjusted Odds Ratio controlling for gender, age and clustering of students within schools (N=848) *

Outcome	Treatment conditions (6 grade only as reference)	6 months post intervention AOR (95% CI)	*p*
Date physical victimization	6, 7 and 8 grades	1.03 (0.61,1.73)	0.92
	6 and 7 grades	1.06 (0.46,2.44)	0.88
Date sexual victimization	6, 7 and 8 grades	0.98 (0.41,2.36)	0.97
	6 and 7 grades	1.22 (0.57,2.59)	0.58
Date physical perpetration	6, 7 and 8 grades	0.66 (0.32,1.39)	0.25
	6 and 7 grades	0.65 (0.21,1.94)	0.41
Date sexual perpetration	6, 7 and 8 grades	0.81 (0.25,2.63)	0.70
	6 and 7 grades	0.7 (0.11,4.34)	0.69

*reference intervention group: 6 grade only or 6 grade longitudinal

Table 6 (see below) covers the frequency of physical dating violence and sexual dating violence at the six-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the 6th and 7th grade group and the 6th, 7th and 8th grade group were statistically significant, suggesting that the treatment effects were equal.

Table 6: Six-month intervention effect on **frequency** of dating violence: Regression coefficients controlling for gender, age, and clustering of students within schools (N=848)*

Outcome	Treatment conditions (6 grade only as reference)	6 months post intervention Regression Coefficient (95% CI)	*p*
Date physical victimization	6, 7 and 8 grades	-0.01 (-0.11,0.09)	0.91
	6 and 7 grades	-0.01 (-0.14,0.13)	0.95
Date sexual victimization	6, 7 and 8 grades	-0.01 (-0.11,0.10)	0.92
	6 and 7 grades	-0.03 (-0.12,0.17)	0.55
Date physical perpetration	6, 7 and 8 grades	-0.03 (-0.11,0.05)	0.44
	6 and 7 grades	-0.05 (-0.14,0.05)	0.30
Date sexual perpetration	6, 7 and 8 grades	-0.03 (-0.11,0.04)	0.36
	6 and 7 grades	-0.03 (-0.11,0.04)	0.37

*reference intervention group: 6 grade only or 6 grade longitudinal

Twelve-month follow-up outcome data. Our twelve-month follow-up data allows us to address the question of whether additional dosages of SB leads to greater reductions in DV/H than single dosages (6th grade longitudinal versus 6th grade only groups) and assess our first saturation-level question on the relative value of multiple grades receiving SB compared to one grade at the 12-month follow-up mark.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

For these analyses we do not combine our 6th grade only group (n= 3 schools with 128 students) with our 6th grade longitudinal group (n= 4 schools with 234 students) because at the 12-month follow-up mark the 6th grade longitudinal group has received the 6th grade intervention when they were in sixth grade and the 7th grade intervention when they were in seventh grade. As the name indicates, the 6th grade only group only receives the intervention in 6th grade and not again in 7th grade. Our third group is made up schools assigned to receive the 6th and 7th grade SB intervention or the 6th, 7th and 8th grade SB intervention (n= 6 schools and 452 students for both sets of schools).

Like the last section, we present our 12-month follow-up results in four tables (Tables 7-10). Each model in each of the four tables includes AOR values (prevalence outcomes) or regression coefficients (frequency outcomes) and their 95% confidence intervals (in parentheses) with p values for each treatment arm, with the 6th grade only group serving as the reference group in each model. Each model includes covariates for gender, age, prior victimization and perpetration (matched to each outcome variables such that if the outcome is 6th month rate of sexual harassment victimization we include a control variable of baseline sexual harassment victimization rate) and a control for clustering of students within schools.

Table 7 (see below) covers the prevalence of peer violence and sexual harassment at the twelve-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the combined 6th and 7th grade group with the 6th, 7th and 8th grade group or the 6th grade longitudinal group were statistically significant, suggesting that the treatment effects were equal.

Table 7: Twelve-month intervention effect on **prevalence** of peer violence and sexual harassment: Adjusted Odds Ratio controlling for gender, age, wave I sexual harassment victimization and perpetration and controlling for clustering of students within schools (N=814)

Outcome	Treatment conditions (6 grade only as reference)	AOR (95% CI)	*p*
Peer physical victimization	6,7 grades or 6-8 grades	0.87 (0.43,1.76)	0.67
	6 grade longitudinal	0.83 (0.48,1.43)	0.45
Peer sexual victimization	6,7 grades or 6-8 grades	1.32 (0.65,2.68)	0.40
	6 grade longitudinal	0.92 (0.40,2.14)	0.83
Peer physical perpetration	6,7 grades or 6-8 grades	1.23 (0.44,3.49)	0.66
	6 grade longitudinal	1.16 (0.48,2.81)	0.71
Peer sexual perpetration	6,7 grades or 6-8 grades	1.04 (0.36,3.01)	0.93
	6 grade longitudinal	0.8 (0.31,2.09)	0.60
Sexual harassment victimization	6,7 grades or 6-8 grades	0.75 (0.39,1.42)	0.33
	6 grade longitudinal	0.86 (0.44,1.67)	0.61
Sexual harassment perpetration	6,7 grades or 6-8 grades	1.51 (0.55,4.17)	0.92
	6 grade longitudinal	1.42 (0.51,3.99)	0.77

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 8 (see below) covers the frequency of peer violence and sexual harassment at the six-month post intervention mark. We had two statistically significant effects (i.e., $p < .05$) in Table 8. The 6th & 7th grade intervention or 6th, 7th & 8th grade interventions were associated unexpectedly with more peer sexual violence perpetration compared to the 6th grade only intervention (Regression coefficient = 0.04, p= 0.04). The standardized effect size for this finding was a Cohen's D of 0.23 which is equivalent to a small effect (Cohen, 1988). Our next statistically significant finding was that the 6th & 7th grade intervention or 6th, 7th & 8th grade intervention were associated with (as we anticipated) less sexual harassment victimization compared to the 6th grade only intervention (Regression coefficient = -0.22, p= 0.03). The standardized effect size for this finding was a Cohen's D of 0.26 which is equivalent to a small effect (Cohen, 1988). We also had one borderline statistically significant finding (p<.10).

Our 6th grade longitudinal group that received two dosages of the SB intervention (one in 6th and one in 7th grade) was associated with (as we anticipated) less sexual harassment victimization compared to the 6th grade only intervention (Regression coefficient = -0.18, p= 0.08). The standardized effect size for this finding was a Cohen's D of 0.22 or a small effect (Cohen, 1988). None of the other treatment comparisons between the reference category of the 6th grade-only group to the other groups were statistically significant.

Table 8: Twelve -month intervention effect on **frequency** of peer violence and sexual harassment: Adjusted Odds Ratio controlling for gender, age, wave I sexual harassment victimization and perpetration and controlling for clustering of students within schools (N=814)

Outcome	Treatment conditions (6 grade only as reference)	6 months post intervention Regression Coefficient (95% CI)	*p*
Peer physical victimization	6,7 grades or 6-8 grades	-0.07 (-0.32,0.17)	0.52
	6 grade longitudinal	-0.13 (-0.34,0.08)	0.19
Peer sexual victimization	6,7 grades or 6-8 grades	0.02 (-0.09,0.12)	0.76
	6 grade longitudinal	-0.06 (-0.15,0.04)	0.21
Peer physical perpetration	6,7 grades or 6-8 grades	0.02 (-0.18,0.22)	0.79
	6 grade longitudinal	-0.01 (-0.12,0.11)	0.88
Peer sexual perpetration	6,7 grades or 6-8 grades	0.04 (0,0.07)	0.04
	6 grade longitudinal	0.02 (-0.02,0.06)	0.21
Sexual harassment victimization	6,7 grades or 6-8 grades	-0.22 (-0.42,-0.02)	0.03
	6 grade longitudinal	-0.18 (-0.38,0.02)	0.08
Sexual harassment perpetration	6,7 grades or 6-8 grades	-0.01 (-0.13,0.10)	0.80
	6 grade longitudinal	0 (-0.11,0.11)	0.99

Table 9 (see below) covers the prevalence of physical dating violence and sexual dating violence at the twelve-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the combined 6th and 7th grade group with the 6th, 7th and 8th grade group or the 6th

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

grade longitudinal group were statistically significant, suggesting that the treatment effects were equal.

Table 9: Twelve -month intervention effect on **prevalence** of dating violence: Adjusted Odds Ratio controlling for gender, age, and clustering of students within schools (N=443)

Outcome	Treatment conditions (6 grade only as reference)	AOR (95% CI)	*p*
Date physical victimization	6,7 grades or 6-8 grades	1.12 (0.39,3.20)	0.81
	6 grade longitudinal	0.6 (0.20,1.81)	0.31
Date sexual victimization	6,7 grades or 6-8 grades	1.02 (0.15,6.76)	0.99
	6 grade longitudinal	0.66 (0.07,6.07)	0.68
Date physical perpetration	6,7 grades or 6-8 grades	1.4 (0.31,6.31)	0.61
	6 grade longitudinal	1.5 (0.33,6.91)	0.54
Date sexual perpetration[#]	6,7 grades or 6-8 grades	NA	NA
	6 grade longitudinal	NA	NA

[#]Model did not converge due to sparse cells

Table 10 (see below) covers the frequency of physical dating violence and sexual dating violence at the twelve-month post intervention mark (including victimization and perpetration). None of the treatment comparisons between the reference category of the 6th grade-only group to the combined 6th and 7th grade group with the 6th, 7th and 8th grade group or the 6th grade longitudinal group were statistically significant, suggesting that the treatment effects were equal.

Table 10: Twelve -month intervention effect on **frequency** of dating violence: Regression coefficients controlling for gender, age, and clustering of students within schools (N=443)

Outcome	Treatment conditions (6 grade only as reference)	6 months post intervention Regression Coefficient (95%CI)	*p*
Date physical victimization	6,7 grades or 6-8 grades	-0.01 (-0.14,0.11)	0.80
	6 grade longitudinal	-0.06 (-0.17,0.06)	0.28
Date sexual victimization	6,7 grades or 6-8 grades	0.01 (-0.09,0.10)	0.89
	6 grade longitudinal	-0.03 (-0.12,0.06)	0.46
Date physical perpetration	6,7 grades or 6-8 grades	-0.01 (-0.19,0.18)	0.94
	6 grade longitudinal	0.01 (-0.18,0.20)	0.92
Date sexual perpetration	6,7 grades or 6-8 grades	0.03 (-0.03,0.08)	0.28
	6 grade longitudinal	0 (-0.05,0.05)	0.82

Discussion

The current study builds on two prior RCTs assessing the Shifting Boundaries program, one of which was conducted in the same study population of NYC middle schools. In our prior multi-level research of the Shifting Boundaries program (NYC-1) — classroom curricula (SBC)

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

and building-wide activities (SBS) — we found that SBS and the SBC+SBS interventions were effective at reducing dating and peer violence victimization and perpetration (Taylor et al., 2013a). That is, the combination of the classroom and building interventions as well as the building intervention alone statistically reduced sexual harassment (victimization and perpetration) by 26-34% six months post follow-up. SBS statistically reduced victimization and perpetration of physical and sexual dating violence by about 50% up to six months after the intervention. SBC+SBS and SBS alone led to 32-47% statistically lower peer sexual violence victimization and perpetration up to six months after the intervention. While the preponderance of results indicates that the interventions were effective in reducing violent incidents, a few anomalous results (e.g., reported declines in total peer violence frequency which were contradicted by higher prevalence estimates) did emerge. However, after careful analysis these anomalous results were deemed to be most likely spurious. The current RCT was designed to assess (1) the effects of saturating a school environment by providing the full SB program to all three middle school grades compared to only two grades or one grade and (2) the effects of two dosages of the SB intervention across two years compared to one dosage of the SB intervention across one year. While we varied the saturation levels and dosages of SB in NYC-2, each treatment group in NYC-2 received the SB classroom curriculum (SBC) and SB school (building-level) (SBS) components.

Few differences across the treatment groups.

Our NYC-2 results indicate that, overall, providing the SB treatment to only one grade level in a middle school does just as well in terms of peer violence and dating violence outcomes as a more saturated process of treating multiple grades in the school. At both the 6-month and the 12-month assessments, however, there was evidence that additional saturation beyond one grade is associated with reductions in sexual harassment victimization. Schools that delivered SB to both 6th and 7th graders (compared to just 6th graders) showed reductions SH victimization reports at 6 months post treatment, an effect that was still significant at the 12-month assessment. Also at 12 months post treatment, schools that delivered SB to all middle school grades showed reductions in self-report of SH victimization.

However, we also found that greater saturation of the SB program (delivered to 6th & 7th graders or to all three grades levels) was unexpectedly associated with more reported perpetration of sexual violence against peers at 12 month post treatment compared to the 6th grade only group. This higher level of reporting of perpetrating sexual violence against peers could mean that the program increased this form of violence and is an iatrogenic finding. That is, we had a true backfire effect and the intervention increased the proportion of students that were perpetrators of peer sexual violence. It is also possible that we might have a reporting effect issue. That is, the intervention sensitized students to recognizing that they are a perpetrator of sexual violence so they are more likely to report this on the survey. Under this interpretation, the intervention is helping students recognize these acts as violence and perhaps increasing the sensitivity of this group of participants to the illicit nature of sexual violence and improving their likelihood of reporting this form of violence.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Another alternative is that it could be a spurious finding because it is not supported by other examples of increases in violent behavior associated with greater saturation of SB. In fact, this finding stands in contrast with the additional borderline statistically significant findings ($p<.10$) at the 6-month assessment suggest that receiving SB dosage for two grades rather than only one was associated with reduced frequency of peer physical victimization frequency and peer sexual violence perpetration.

Overall, future research could potentially benefit from including a measure on student understanding of key project concepts. Such a measures might aid the research team in assessing whether this finding was a result of a spurious process or true student reflection of their actions and understanding that they perpetrated sexual violence. Determining if students have mastered the learning objectives of the SB program may contribute to the validity of student self-report and provide some reasoning for increases in reporting. In other words, such a measure might help us sort out the extent to which student knowledge of the content and skills of the SB program might contribute to the student's ability to report accurately and effect increases in initial reporting.

There were no results indicating that offering the SB program to a grade of students in two successive years (the 6th grade longitudinal design) resulted in statistically differential effects compared to a one-time delivery of the program in 6th grade. This finding is analogous to Foshee and colleagues' study that found no additional TDV reductions associated with a booster session delivered two years (our intervention was one-year later) after the original application of the intervention (Foshee, Bauman, et al., 2004). While Foshee et al. did not find any additional benefits of additional later interventions, additional booster effects have been detected with other adolescent problem behavior interventions related to smoking prevention (Botvin, Renick, & Baker, 1983; Dijkstra, Mesters, De Vries, Van Breukelen, & Parcel, 1999) and substance use (Botvin, Baker, Filazzola, & Botvin, 1990; Spoth, Redmond, & Shin, 2001). One borderline statistically significant effect in the small effect size range (SB program delivered to 6th graders in year 1 and again to the same students, as 7th graders, in year 2 was associated with less sexual harassment victimization frequency compared to the 6th grade only intervention) highlights the potential potency of multiple dosages of the SB program for sexual harassment prevention work. Nevertheless, in sum, the weight of evidence suggests that increased SB program dosage over the middle school years is not indicated to address peer and dating violence and sexual harassment outcomes.

From a policy and administrative perspective, these results largely support a minimalistic approach, in that SB effectiveness for peer and dating violence outcomes may be achieved by delivery to only one grade level in middle schools. However, taking these results in the context of our earlier work (NYC-1), there is a rationale for considering saturated delivery of the SBS component. In earlier research, SBS was effective at reducing DV/H outcome independent of the classroom curriculum (SBC). Because the SBS program can be introduced to an entire middle school at low-cost, and our current research shows positive effects of exposing more than just a single grade to the SB program, these results taken together suggest policy and administrative consideration of a saturated delivery of the SBS program. What remains to be

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

examined in the study of the SBS intervention is protocols to support school maintenance of SBS activities beyond the research phase and low-cost methods for ongoing program assessment internal to school administrations. For long-term use, schools will be able to look beyond our short-term needs of a study and introduce a variety of time saving efficiencies. For example, recent advances in mapping software can be used to create e-maps of the school floorplan to save student and administrative time in completing the mapping exercise, improve the mapping precision, and ease of tallying results to inform administrative responses. Also, the respecting boundary agreements could be digitized and integrated into a database to facilitate data collection by school staff and tracking of cases of DV/H.

Our project also added a grade-differentiated curricula, the Safe Dates curriculum for 8th graders, as a component to our test of this SB program experiment. In our experiment, we did not find strong evidence for the inclusion of additional grade levels of prevention curricula, whether the SB material for 7th graders or the combined SB and Safe Dates material for 8th graders. Despite the positive results regarding Safe Dates, drawn on implementation of the program in one location (rural North Carolina) (Foshee et al., 1998; Foshee, Bauman, et al., 2004; Foshee, Bauman, et al., 2005; Foshee et al., 2000), Safe Dates did not seem to lead to additional DV/H reductions when combined with SB. However, we only included four lessons from Safe Dates for the 8th grade students (plus three lessons from SB) and did not attempt to isolate the effects of the Safe Dates curriculum.

Resistance from schools and its effects on study.

A key issue to emerge in this study was the significant resistance from schools to participation in the study. All accounts seem to suggest that it was not something about the SB program that led to problems but a general resistance to any prevention or gender violence interventions in the NYC schools at the time of our study. As outlined earlier (see 'Challenges in Implementing our Experiment' in the Methods section), there were significant policy changes occurring in the NYC schools during the time of the study, a major change in the school administration leadership over the study (and a new mayor along with a new chancellor), reorganization of the NYC school department at the beginning of the new mayor's term in January 2014, drastic cutbacks in staffing at the NYC schools removing the necessary personnel to implement the SB intervention or even leading to schools being closed, and Hurricane Sandy prevented a number of schools from continuing their participation.

All of these barriers required our team to make a number of changes in our planned design. First, we had to drop the no treatment control group, with a smaller group of schools willing to participate in the study we had to focus on the main research question of the comparative effectiveness of different levels of SB treatment, the earlier NYC experiment a couple of years before already addressed the question of treatment efficacy (treatment versus no-treatment). While we have confidence in our earlier findings of SB being effective at reducing DV/H compared to no treatment (Taylor et al., 2013a), it would have been useful to replicate those results and confirm that where we found that additional saturation beyond one grade is associated with reductions in sexual harassment victimization (at both 6 and 12-month follow-up) that the groups with the additional saturation were also better at reducing sexual

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

harassment victimization than a control group. It is a limitation of this NYC-2 study that we cannot confirm that finding with a control group.

Also, with the lower than expected levels of school participation, we had to drop our plan to have a balanced design with the same number of schools in each study condition. That is, after we experienced a large drop off of schools we had to randomize new schools into the study as they emerged ('randomization on the fly' as opposed to all at once like in the NYC-1 study), leading to an imbalanced number of schools in each study condition. Despite this variation in sample sizes across our four comparison groups, our power analyses revealed that we still had adequate statistical power (see power analysis section) to detect differences across our comparison groups (i.e., we could find small and medium sized effects for some of our analyses) for our six month analyses but less power for our 12-month follow-up analyses (i.e., we could find medium to large sized effects).

Next, we hoped to have our 6th grade longitudinal group receive three years of treatment. However, too few of the schools were willing to continue participation beyond two years of treatment due to competing academic demands within the schools and staffing deficits leading to a resistance to applying limited resources to DV/H prevention. Similarly, we had hoped to have up to a 24-month follow-up survey but that plan had to be abandoned given the lack of willingness of schools to continue participation. The schools agreed to one-year of follow-up surveys. Nevertheless, the one-year follow-up still represented a longer follow-up period than the earlier research on SB that only had a six-month follow-up period (Taylor et al., 2008; Taylor et al., 2013a; Taylor, 2010). Despite these problems, we continued the project and in the end we believe some important findings emerged and we learned some important lessons about navigating a field experiment through multiple administrative land mines and an actual hurricane.

Limitations

The general limitations of self-reports are applicable to this study. For example, students may not be able to recall the timing of a violent act or may have deliberately under-reported certain behavior (e.g., they may not want to admit they were victimized) (Jackson et al., 2000), or may have exaggerated certain behavior (e.g., over-reported the number of times they were physically abusive with a girl). Despite these potential problems, which likely were balanced across our comparison groups, self-report surveys (especially confidential surveys like the type used in the study) have become an accepted modality of collecting data on the subject matter of DV/H and violence more generally.

Another key limitation associated with self-reports is that they are a limited modality in terms of capturing the intensity and context of violent behavior and harassment (Wolfe et al., 2009). Like other researchers in this area, we measured DV/H by having participants answer questions on whether they have performed a specific act against a partner or peer, such as pushing, kicking, hitting, etc. (or been the victim of these acts). These types of reports do not encompass motivations or circumstances surrounding violent acts or distinguish between acts of offense or defense (Wolfe et al., 2009). The 'victimization' experience of being physically

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

attacked could be the result of an unprovoked attack from a perpetrator or someone who started a fight with a stronger person and ended up getting beat up in the process. Therefore, as Foshee and colleagues observed (Foshee et al., 2014), program effects on victimization and perpetration may not reflect effects on "victimization" and "perpetration" per se but rather effects on producing a less violent environment in general.

While some past DV/H studies have not included measures of sexual victimization (Wolfe et al., 2009), our study did have such measures. Nevertheless, because of concerns raised by school personnel on the sensitivity of such questions for a middle school population, as in our NYC-1 study (Taylor et al., 2013a), we were limited in how we could measure sexual victimization to two main items ("pushed, grabbed, shoved, or kicked you in your private parts" and "made you touch their private parts or touched yours when you did not want them to"). However, Foshee et al. used only two items to measure sexual violence in her evaluation of Safe Dates (Foshee et al., 1998).

Also, to maintain continuity with our earlier research in Cleveland and NYC-1 we used the same measures. One drawback of that approach was that we did not have any of the updated items on dating violence or sexual harassment perpetrated via the web (e.g., social media) or by other electronic means (e.g., texting). Despite the high rates of ARA/SH reported in our study, it is possible this led to an underestimation of ARA or SH in our study.

Our study was limited to two follow-up (six and twelve months post-treatment) and it is unclear whether our findings would change over a longer follow-up period. For example, Foshee and colleagues (Foshee et al., 2000; Foshee, Benefield, et al., 2004; Foshee et al., 1996a) conducted longitudinal follow-up surveys at one year and four years following their intervention (Safe Dates) and found that come of the behavioral effects evident at one-month follow-up dissipated after four years post-program. With a sample of 9th grade students in Ontario, Canada, Wolfe and colleagues (Wolfe et al., 2009) found significant effects 2.5 years post treatment (a 21-lesson curriculum delivered by teachers with additional training in the dynamics of dating violence and healthy relationships) for physical dating violence. As with other violence prevention studies with limited longitudinal data (Ackard et al., 2007; Chiodo et al., 2009; Foshee, Ennett, Bauman, & Suchindran, 2005; O'Leary & Slep, 2003; Roberts et al., 2003) longer follow-up is important. With longer follow-up and large samples, researchers can examine differences in trajectories of youth violence by various age and gender subgroups.

Another measurement limitation is that we did not ask participants about their sexual orientation, so it is not possible to determine if our findings would vary for same sex relationships. We also were not able to measure some important covariates (e.g., violence in the home or community) which may have potentially influenced our findings. However, given our use of a randomized experiment, these unmeasured variables should have (by design) been balanced across the comparison groups.

While similar DV/H studies were more limited in their applicability to different ethnic groups — the Fourth R research sample was mainly White youth (Wolfe et al., 2009) and the

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Safe Dates research sample was mainly rural White youth (Foshee et al., 1998) — our sample included a broader range of ethnic groups (see descriptive statistics section). However, our study was also done in the largest school district in the U.S. (NYC), and our results are possibly only applicable to similar very large urban districts.

There are several threats to the validity of our experiment that we reviewed in the Methods section. Our overall conclusion was that our experimental design achieved its basic purpose of creating comparable conditions to assess outcome differences in our comparison groups. While we found some differences between the comparison conditions prior to the experiment (during the baseline period), most of these differences (while statistically significant) were not very large differences. For the most part, the four study groups/conditions were similar on the majority of our measures leaving the only major differences across the groups their assigned intervention condition. Additionally, random assignment procedures were followed closely (no "overrides"). Despite schools dropping out of the study, the schools that stayed in the study adhered to their assigned treatment. Assessing the interest and wherewithal of schools to stick through a field experiment is difficult enough. However, once schools started dropping out, we had to try to implement the experiment with schools that expressed only a peripheral interest in the project which led to more dropouts. Finally, we included the variables where there were pre-treatment differences into our outcome models as covariates to remove any potential biases these small imbalances might have presented for the interpretation of our results.

Another major concern in our study was whether attrition in our study created any pattern of bias that would interfere with our ability to draw unequivocal inferences from our study (see "Attrition analyses"). Overall, we did not observe much by way of patterns in our study for the schools that continued on to complete the follow-up survey waves and those schools that dropped out after doing only a baseline survey. We observed few differences between the dropout schools and the completer schools on a variety of background factors and violence measures. Where there were some differences, we addressed this in our statistical modeling. For example, for the survey data, we found differences by age and the level of pre-treatment exposure to sexual harassment for our fully participating schools compared to the dropout schools. To address this issue, in our outcome models we include, among other variables, age and pre-treatment exposure to sexual harassment as covariates. Therefore, whatever impact these small differences might have on our outcome models are controlled for through the use of covariates.

Implications

NYC-1 results indicated that the SB program was effective at reducing various forms of sexual harassment/gender violence relative to a no-treatment group. The overall result from NYC-2 (that contained the SBC and SBS components) is that for most of our measures, providing the SB program to only one grade level in a middle school does just as well in terms of peer violence and dating violence outcomes as a more saturated process of treating multiple grades in the school. The same applies for multiple dosages across two years for one grade cohort.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Considering the full set of results from the NYC-1 and NYC-2 studies, we have strong evidence of the effectiveness of the school-wide program (SBS) at reducing DV/H outcomes independent of the classroom curriculum (Taylor et al., 2013a) and our current research shows some positive effects of exposing more than just a single grade to the SB program (e.g., reducing sexual harassment). Because SBS can be introduced to an entire middle school at low-cost, these results taken together suggest that schools should consider a saturated delivery of the SBS intervention. Holding to a basic minimalist recommendation to conserve resources, we believe that our findings suggest that school administrators should focus on the SBS component of the SB program that was found to be successful in NYC-1 and offer SBS in a saturated manner across all three grades. That is, building on our NYC-2 result (that additional saturation beyond one grade is associated with reductions in sexual harassment victimization at 6 and 12-months follow-up) and the logistical straightforwardness of implementing SBS across the entire school environment (for little extra cost beyond just the 6th grade), we believe there is empirical justification for such an approach.

Another implication from the NYC-2 results is that more evaluation work is needed to explore the benefits of a saturated delivery of SBS across all grades or additional dosages of SBS. First, the above described barriers precluded our team from having a no-treatment control group. We want to have assurance that the different dosage and saturation levels are all at least better than doing no intervention. Our NYC-2 interpretations rely on an assumption we make based on the NYC-1 results conducted with a similar sample, location and general time frame. Future studies will want to have a control group so that if the researchers find that more saturation is better than less saturation that they can be assured that more saturation is also better than no intervention. We also had modest statistical power in our study. While we had enough statistical power to find small effects for certain comparisons at the six month follow-up, for other comparisons we could only detect medium or large effect sizes. Given the developmental nature of SB it is more reasonable to find more small effects and the RCT should be powered with a larger number of schools ($n \geq 10$ schools per treatment arm) to detect such effects. Also, given that SB has only been evaluated in two communities (suburban Ohio, outside of Cleveland and large urban NYC), it is still necessary to assess whether our results with SB can be replicated in other communities, including rural areas.

This study was conducted to address the serious problem of youth DV/H through the testing of a prevention program intervention for students in middle school. Most research in this area had focused on older middle and high school students, whereas we believed those groups were less appropriate as a primary prevention audience. We found that a relatively large number of middle school students had experienced dating violence (20% reported having been the victim of any physical dating violence, 59% reported having been the victim of any physical peer violence and 49% reported having been sexually harassed). Our lukewarm findings regarding the effectiveness of saturating the middle school environment or providing additional dosages suggests that the field may need to work with even younger groups to invoke a true primary prevention effort. However, we are not aware of much work being done with elementary school students in the area of the primary prevention of youth sexual harassment/ relationship gender violence. Since our intervention is designed for middle school students, our material

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

would have to be adapted to be developmentally appropriate for elementary school students or new interventions would have to be designed. However, this is another age group to look to for preventing or at least reducing DV/H.

We also believe more work is needed to better understand the mechanisms of changing behavior through prevention interventions. Future research in this area would benefit from the collection of additional survey variables that would explain the change process. Our project was constrained to a short time frame (about 30 minutes) to collect our survey data and had to focus mainly on our outcome measures for the survey items. Future research might include looking at those attributes that contribute to prevention of peer and dating violence and sexual harassment (e.g., empathy, respecting others, and handling emotions). Such research might also help determine if providing instruction in elementary school with a focus on these attributes would lead to reduction in peer violence as well as effects on a K-12 scaffolder approach.

Also, we believe that phenomenological interviews with student participants would allow researchers to explore these mechanisms in a rigorous manner. Phenomenological studies are highly systematic qualitative methods for inquiry and analysis (Creswell, 1998; Patton, 1990) and allow the researcher to enter the field of perception of the program participants (Creswell, 1998) to elucidate what essential program experience the students described that caused them to change or not change. In the context of studying DV/H prevention programs, a phenomenological interviewing approach is especially warranted given the dearth of knowledge on how students do and do not change; what they learn or fail to learn in treatment; what they respond or relate to most about treatment; and non-treatment factors that may lead to change in behaviors. These qualitative interviews would give student participants an opportunity to articulate any changes they felt they had experienced as a result of the program in their own words. These qualitative data would shed light on how or why change did or did not take place, which components of the program they believe contributed the most to that change, and describe what factors outside of treatment (e.g., positive and negative peer and/or teacher support) may be meaningful to that change.

Next, given the problems we experienced with recruiting schools to our experiment and the large number of schools that dropped out from our study, a case can be made that investing resources in research to address these methodological concerns would be very useful for the field. Such methodological research could lead to more rigorous research, completed in more timely fashion with fewer drained resources. Additional planning resources to handle the various contingencies that emerged in this project could have been useful. Also, having documents like this final report that chronicle what happened in this project and how the barriers were addressed will help future researchers if they run into similar problems. Federal agencies that support research could help convene researchers to draw out the lessons learned from these types of field evaluations and make sure the lessons learned enter the lexicon of other researchers these agencies fund or potentially fund. More work is needed on best practices in how to resolve these types of problems or future research in this important area will be plagued by similar problems.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Conclusion

A sizeable literature has emerged that documents the significant negative effects of youth DV/H, whether experienced as a *victim* (Banyard & Cross, 2008; Callahan et al., 2003; Chiodo et al., 2012b; Exner-Cortens et al., 2013; Holt & Espelage, 2005; Howard & Wang, 2003a; Howard & Wang, 2003b; McDonald et al., 2010) or *perpetrator* (Calvete, Orue, Gamez-Guadix, & de Arroyabe, 2014; Johnson, Giordano, Longmore, & Manning, 2014; Nahapetyan, Orpinas, Song, & Holland, 2014). Middle school present a critical opportunity for addressing DV/H since virtually the entire population passes through them during this critical period of adolescence. With some school districts requiring use of violence prevention interventions at the same time as school burdens are rising with high stakes testing and decreasing resources, we recommend implementation of SBS by school administrators and counseling staff.

Despite the difficult path it took to complete NYC-2, and the associated changes we had to make to the study (dropping our control group, reducing the dosage testing to two instead of three dosages, and reducing the follow-up period to 1 year instead of potentially 3 years), we still believe some important new knowledge emerged from the NYC-2 study. In addition to lessons learned about how to navigate a field experiment through difficult administrative and resource barriers (see section on "Challenges in Implementing our Experiment"), we learned after NYC-2 that for the most part additional dosages and saturation do not alter the findings for most of our outcome measures compared to just implementing the SB program with just 6th grade students. Providing the SB treatment once in 6th grade works as well in terms of DV/H levels as applying it once per year for two years with the same group (in 6th and 7th grades). Likewise, implementing SB with only one grade level in a middle school does just as well in terms of <u>peer violence and dating violence</u> outcomes as a more saturated process of treating multiple grades in the school. However, we did find that additional saturation beyond one grade is associated with reductions in <u>sexual harassment</u> victimization at the 6 and 12-month follow-up period. Considering our NYC-1 and NYC-2 results together, we believe there is empirical justification for implementing the school-wide component SBS across the entire school environment. That is, we feel the data support implementing the SB program for at least the 6th grade students but given the nature of the SBS intervention that can be extended to the whole school with little extra cost.

As a result of this and prior studies, a body of scientific data is emerging about the beneficial effects of DV/H interventions targeted to middle school students. We encourage other researchers and program developers to expand on this study as they pursue efforts to interrupt the precursors to youth dating violence and sexual harassment.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

References

AAUW Educational Foundation. (1993). *Hostile hallways: The AAUW survey on sexual harassment in America's schools*. Washington, DC: Author.

AAUW Educational Foundation. (2001). *Hostile hallways: Bullying, Teasing, and Sexual Harassment in School*. Washington, DC: Author.

Ackard, D. M., Eisenberg, M. E., & Neumark-Sztainer, D. (2007). Long-term impact of adolescent dating violence on the behavioral and psychological health of male and female youth. *Journal of Pediatrics, 151*(5), 476-481. doi: 10.1016/j.jpeds.2007.04.034

Ackard, D. M., & Neumark-Sztainer, D. (2002). Date violence and date rape among adolescents: associations with disordered eating behaviors and psychological health. *Child Abuse & Neglect, 26*(5), 455-473.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Ajzen, I., & Fishbein, M. (Eds.). (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

American Association of University Women, A. (1993). *Hostile hallways: The AAUW survey on sexual harassment in America's schools.* Washington, DC: Author.

American Association of University Women, A. (2001). *Hostile hallways: Bullying, Teasing, and Sexual Harassment in School*. Washington, DC: Author.

Arriaga, X. B., & Foshee, V. A. (2004). Adolescent dating violence - Do adolescents follow in their friends', or their parents', footsteps? *Journal of Interpersonal Violence, 19*(2), 162-184. doi: 10.1177/0886260503260247

Astor, R. A., Meyer, H. A., & Pitner, R. O. (2001). Elementary and middle school students' perceptions of violence-prone school subcontexts. . *The Elementary School Journal, 101*(5), 511-528.

Astor, R. A., Meyer, H. M., & Behre, W. J. (1999). Unowned places and times: Maps and interviews about violence in high schools. . *American Educational Research Journal, 36*(1), 3-42.

AveryLeaf, S., Cascardi, M., Oleary, K. D., & Cano, A. (1997). Efficacy of a dating violence prevention program on attitudes justifying aggression. *Journal of Adolescent Health, 21*(1), 11-17.

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research, 20*(1), 40-49.

Banyard, V. L., & Cross, C. (2008). Consequences of teen dating violence - Understanding intervening variables in ecological context. *Violence against Women, 14*(9), 998-1013. doi: 10.1177/1077801208322058

Basile, K. C., Espelage, D. L., Rivers, I., McMahon, P. M., & Simon, T. R. (2009). The theoretical and empirical links between bullying behavior and male sexual violence perpetration. *Aggression and Violent Behavior, 14*(5), 336-347.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Bentley, C. G., Galliher, R. V., & Ferguson, T. J. (2007). Associations among aspects of interpersonal power and relationship functioning in adolescent romantic couples. *Sex Roles, 57*(7-8), 483-495. doi: 10.1007/s11199-007-9280-7

Bergman, L. (1992). Dating Violence Among High-School-Students. *Social Work, 37*(1), 21-27.

Berk, R. A., Boruch, R. F., Chambers, D. L., Rossi, P. H., & Witte, A. D. (1985). Social-Policy Experimentation - a Position Paper. *Evaluation Review, 9*(4), 387-429.

Berk, R. A., Smyth, G. K., & Sherman, L. W. (1988). When random assignment fails: Some lessons from the Minneapolis Spouse Abuse Experiment. . *Journal of Quantitative Criminology, 4*, 209-223.

Berkowitz, A. D. (2010). Fostering healthy norms to prevent violence and abuse: The social norms approach. *The prevention of sexual violence: A practitioner's sourcebook*, 147-171.

Boruch, R. (1997). *Randomized Experiments for Planning and Evaluation: A Practical Guide*. Thousand Oaks: SAGE.

Boruch, R. F., McSweeny, A. J., & Soderstrom, E. J. (1978). Randomized Field Experiments for Program Planning, Development, and Evaluation - Illustrative Bibliography. *Evaluation Quarterly, 2*(4), 655-695.

Boruch, R. F., & Wolhke, W. (Eds.). (1985). *Randomization and field experimentation.* San Francisco: Jossey-Bass.

Botvin, G. J., Baker, E., Filazzola, A. D., & Botvin, E. M. (1990). A cognitive-behavioral approach to substance abuse prevention: One-year follow-up. *Addictive Behaviors, 15*(1), 47-63.

Botvin, G. J., Renick, N. L., & Baker, E. (1983). The effects of scheduling format and booster sessions on a broad-spectrum psychosocial approach to smoking prevention. *Journal of Behavioral Medicine, 6*(4), 359-379.

Braga, A. A., Welsh, B. C., Papachristos, A. V., Schnell, C., & Grossman, L. (2014). The growth of randomized experiments in policing: the vital few and the salience of mentoring. *Journal of Experimental Criminology, 10*(1), 1-28.

Budd, R., Bleiker, S., & Spencer, C. (1983). Exploring the Use and Non-Use of Marijuana as Reasoned Actions - an Application of Fishbein and Ajzen Methodology. *Drug and Alcohol Dependence, 11*(2), 217-224.

Burcky, W., Reuterman, N., & Kopsky, S. (1988). Dating violence among high school students. *The School Counselor, 35*, 353-358.

Callahan, M. R., Tolman, R. M., & Saunders, D. G. (2003). Adolescent dating violence victimization and psychological well-being. *Journal of Adolescent Research, 18*(6), 664-681. doi: 10.1177/0743558403254784

Calvete, E., Orue, I., Gamez-Guadix, M., & de Arroyabe, E. L. (2014). Social Information Processing in Dating Conflicts Reciprocal Relationships With Dating Aggression in a One-Year Prospective Study. *Journal of Interpersonal Violence*, 0886260514564160.

Campbell, D. T. (1969). Reforms as Experiments. *American Psychologist, 24*(4), 409-429.

Campbell, D. T., & Stanley, J. S. (1963). *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.

Centers for Disease Control and Prevention (Producer). (2012a, 3/28/13). Healthy People 2020 Topics & Objectives: Injury and Violence Prevention. Retrieved from http://healthypeople.gov/2020/topicsobjectives2020/overview.aspx?topicid=24

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Centers for Disease Control and Prevention (Producer). (2012b, 3/28/13). Injury Prevention & Control, Violence Prevention: Teen Dating Violence. Retrieved from http://www.cdc.gov/violenceprevention/intimatepartnerviolence/teen_dating_violence.html

Centers for Disease Control and Prevention (Producer). (2014, March 19, 2014). Understanding Teen Dating Violence Fact Sheet. Retrieved from http://www.cdc.gov/violenceprevention/pdf/teen-dating-violence-2014-a.pdf

Centers for Disease Control and Prevention (Producer). (Accessed 2/11/11, 2/11/11). Understanding Teen Dating Violence - Fact Sheet 2010. Retrieved from http://www.cdc.gov/violenceprevention/pdf/TeenDatingViolence_2010-a.pdf

Centers for Disease Control and Prevention, C. (1998). *Measuring Violence-Related Attitudes, Beliefs, and Behaviors Among Youths: A Compendium of Assessment Tools*. Atlanta: CDC, National Center for Injury Prevention and Control.

Chalk, R., King, P. A., & Eds. (1998). *Violence in families: Assessing prevention and treatment programs*. Washington DC: National Academy Press.

Champion, H., Foley, K. L., Sigmon-Smith, K., Sutfin, E. L., & DuRant, R. H. (2008). Contextual Factors and Health Risk Behaviors Associated with Date Fighting Among High School Students. *Women & Health, 47*(3), 1-22.

Chiodo, D., Crooks, C., Wolfe, D., McIsaac, C., Hughes, R., & Jaffe, P. G. (2012a). Longitudinal prediction and concurrent functioning of adolescent girls demonstrating various profiles of dating violence and victimization. *Prevention Science, 13*, 350-359.

Chiodo, D., Crooks, C. V., Wolfe, D. A., McIsaac, C., Hughes, R., & Jaffe, P. G. (2012b). Longitudinal prediction and concurrent functioning of adolescent girls demonstrating various profiles of dating violence and victimization. *Prevention Science, 13*(4), 350-359.

Chiodo, D., Wolfe, D. A., Crooks, C., Hughes, R., & Jaffe, P. (2009). Impact of Sexual Harassment Victimization by Peers on Subsequent Adolescent Victimization and Adjustment: A Longitudinal Study. *Journal of Adolescent Health, 45*(3), 246-252. doi: 10.1016/j.jadohealth.2009.01.006

Cissner, A. B., & Ayoub, L. H. (2014). Building healthy teen relationships: An evaluation of the Fourth R Curriculum with middle school students in the Bronx: U.S. Department of Justice.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coker, A. L., Cook-Craig, P. G., Williams, C. M., Fisher, B. S., Clear, E. R., Garcia, L. S., & Hegge, L. M. (2011). Evaluation of Green Dot: An active bystander intervention to reduce sexual violence on college campuses. *Violence against Women*, 1077801211410264.

Coker, A. L., Fisher, B. S., Bush, H. M., Swan, S. C., Williams, C. M., Clear, E. R., & DeGue, S. (2014). Evaluation of the Green Dot Bystander Intervention to Reduce Interpersonal Violence Among College Students Across Three Campuses. *Violence against Women*. doi: 10.1177/1077801214545284

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Coker, A. L., McKeown, R. E., Sanderson, M., Davis, K. E., Valois, R. F., & Huebner, E. S. (2000). Severe dating violence and quality of life among South Carolina high school students. *American Journal of Preventive Medicine, 19*(4), 220-227.

Conner, M., & Sherlock, K. (1998). Psychosocial determinants of ecstasy use in young people in the UK. *British Journal of Health Psychology, 3*, 295-317.

Conner, R. E. (1977). Selecting a control group: An analysis of the randomization process in twelve social reform programs. *Evaluation Quarterly, 1*, 195-244.

Copp, J. E., & Johnson, W. L. (2015). Patterns, Precursors, and Consequences of Teen Dating Violence: Analyzing Gendered and Generic Pathway. Final Report to the National Institute of Justice. Washington, DC. .

Cornelius, T. L., & Resseguie, N. (2006). Primary and secondary prevention programs for dating violence: A review of the literature. *Aggression and Violent Behavior, 12*(3), 364-375. doi: DOI: 10.1016/j.avb.2006.09.006

Cornfield, J. (1978). Randomization by group: a formal analysis. *American Journal of Epidemiology, 108*, 100-102.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions.* . Thousand Oaks, CA: Sage.

Davis, R. C., & Taylor, B. G. (1995). *Randomized Experiments in a Criminal Justice Setting*. Paper presented at the NIJ Annual Conference on Criminal Justice Research and Evaluation, Washington, DC.

Davis, R. C., & Taylor, B. G. (1997). A proactive response to family violence: The results of a randomized experiment. *Criminology, 35*(2), 307-333.

Davis v. Monroe County Board of Education, 526 U.S. 629 (Supreme Court 1999).

De Vroome, E. M. M., Stroebe, W., Sandfort, R. G. M., De Wit, J. B. F., & Van Griensven, G. J. P. (2000). Safer sex in social context: Individualistic and relational determinants of AIDS-preventive behavior among gay men. *Journal of Applied Social Psychology, 30*(11), 2322-2340.

DeGue, S., Valle, L. A., Holt, M. K., Massetti, G. M., Matjasko, J. L., & Tharp, A. T. (2014). A Systematic Review of Primary Prevention Strategies for Sexual Violence Perpetration. *Aggression and Violent Behavior*.

Dennis, M. L., & Boruch, R. F. (1989). Randomized Experiments for Planning and Testing Projects in Developing-Countries - Threshold Conditions. *Evaluation Review, 13*(3), 292-309.

Dijkstra, M., Mesters, I., De Vries, H., Van Breukelen, G., & Parcel, G. (1999). Effectiveness of a social influence approach and boosters to smoking prevention. *Health Education Research, 14*(6), 791-802.

DuRant, R. H., Altman, D., Wolfson, M., Barkin, S., Kreiter, S., & Krowchuk, D. (2000). Exposure to violence and victimization, depression, substance use, and the use of violence by young adolescents. *Journal of Pediatrics, 137*(5), 707-713.

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., & Hawkins, J. (2010). Youth Risk Behavior Surveillance - United States, 2009. *Morbidity and Mortality Weekly Report, 59*(SS5, Suppl. S), 1-142.

Ebesutani, C., Bernstein, A., Martinez, J. I., Chorpita, B. F., & Weisz, J. R. (2011). The youth self report: applicability and validity across younger and older youths. *Journal of Clinical Child & Adolescent Psychology, 40*(2), 338-346.

Eckstein, R., & Edwards, K.M. . (2014). *Creating a bystander intervention curriculum for the high school environment: Implications for researcher-practitioner collaborations.* . Paper presented at the

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Powerful Partnerships: 20 Years of the Violence Against Women Act and the Path Ahead Meeting, Newark, DE (April, 2014).

Espelage, D. L., & Holt, M. K. (2007). Dating violence & sexual harassment across the bully-victim continuum among middle and high school students. *Journal of Youth and Adolescence, 36*(6), 799-811. doi: 10.1007/s10964-006-9109-7

Exner-Cortens, D., Eckenrode, J., & Rothman, E. (2013). Longitudinal Associations Between Teen Dating Violence Victimization and Adverse Health Outcomes. *Pediatrics, 131*(1), 71-78. doi: 10.1542/peds.2012-1029

Exner-Cortens D., E. J., Rothman E.,. (2012). *Longitudinal Associations Between Teen Dating Violence Victimization and Adverse Health Outcomes*.

Fichman, M., & Cummings, J. N. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods, 6*(3), 282-308.

Fineran, S., & Bennett, L. (1999). Gender and power issues of peer sexual harassment among teenagers. *Journal of Interpersonal Violence, 14*(6), 626-641.

Fineran, S., & Bolen, R. M. (2006). Risk factors for peer sexual harassment in schools. *Journal of Interpersonal Violence, 21*(9), 1169-1190. doi: 10.1177/0886260506290422

Fishbein, M. (1967). Readings in attitude theory and measurement. In K. Glanz , B. K. Rimer, & K. Viswanath (Eds.), *Health behavior and health education: Theory, research, and practice*. San Francisco: Jossey-Bass

Fores, E., Tschann, J. M., & Marin, B. V. (2002). Latina Adolescents: Predicting Intentions to Have Sex. *Adolescence Magazine,* 22-32.

Foshee, V., & Langwick, S. (2010). *Safe Dates: An Adolescent Dating Abuse Prevention CurriculumManual (2nd Edition)*. Center City, MN: Hazelden.

Foshee, V. A. (1996). Gender differences in adolescent dating abuse prevalence, types and injuries. *Health Education Research, 11*(3), 275-286.

Foshee, V. A., Bauman, K. E., Arriaga, X. B., Helms, R. W., Koch, G. G., & Linder, G. F. (1998). An evaluation of Safe Dates, an adolescent dating violence prevention program. *American Journal of Public Health, 88*(1), 45-50.

Foshee, V. A., Bauman, K. E., Ennett, S. T., Linder, F., Benefield, T., & Suchindran, C. (2004). Assessing the long-term effects of the safe dates program and a booster in preventing and reducing adolescent dating violence victimization and perpetration. *American Journal of Public Health, 94*(4), 619-624.

Foshee, V. A., Bauman, K. E., Ennett, S. T., Linder, G. F., Benefield, T., & Schindran, C. (2004b). Assessing the long-term effects of the Safe Dates Program and a booster in preventing and reducing adolescent dating violence victimization and preparation. *American Journal of Public Health, 90*(4), 619-624.

Foshee, V. A., Bauman, K. E., Ennett, S. T., Suchindran, C., Benefield, T., & Linder, G. F. (2005). Assessing the effects of the dating violence prevention program "Safe dates" using random coefficient regression modeling. *Prevention Science, 6*(3), 245-258.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Foshee, V. A., Bauman, K. E., Greene, W. F., Koch, G. G., Linder, G. F., & MacDougall, J. E. (2000). The Safe Dates program: 1-year follow-up results. *American Journal of Public Health, 90*(10), 1619-1622.

Foshee, V. A., Bauman, K. E., Linder, F., Rice, J., & Wilcher, R. (2007). Typologies of adolescent dating violence - Identifying typologies of adolescent dating violence perpetration. *Journal of Interpersonal Violence, 22*(5), 498-519. doi: 10.1177/0886260506298829

Foshee, V. A., Benefield, T. S., Ennett, S. T., Bauman, K. E., & Suchindran, C. (2004). Longitudinal predictors of serious physical and sexual dating violence victimization during adolescence. *Preventive Medicine, 39*(5), 1007-1016. doi: 10.1016/j.ypmed.2004.04.014

Foshee, V. A., Ennett, S. T., Bauman, K. E., & Suchindran, C. (2005). The association between family violence and adolescent dating violence onset does it vary by race, socioeconomic status, and family structure? *Journal of Early Adolescence, 25*(3), 317-344. doi: 10.1177/0272431605277307

Foshee, V. A., Linder, F., MacDougall, J. E., & Bangdiwala, S. (2001). Gender differences in the longitudinal predictors of adolescent dating violence. *Preventive Medicine, 32*(2), 128-141. doi: 10.1006/pmed.2000.0793

Foshee, V. A., Linder, G. F., Bauman, K. E., Langwick, S. A., Arriaga, X. B., & Heath, J. L. (1996a). The Safe Dates project: Theoretical basis, evaluation design, and selected baseline findings. *American Journal of Preventive Medicine, 12*, 39-47.

Foshee, V. A., Linder, G. F., Bauman, K. E., Langwick, S. A., Arriaga, X. B., & Heath, J. L. (1996b). The Safe Dates project: Theoretical basis, evaluation design, and selected baseline findings. *American Journal of Public Health, 81*(1), 45-50.

Foshee, V. A., Linder, G. F., Bauman, K. E., Langwick, S. A., Arriaga, X. B., Heath, J. L., . . . Bangdiwala, S. (1996). The Safe Dates project: Theoretical basis, evaluation design, and selected baseline findings. *American Journal of Preventive Medicine, 12*(5), 39-47.

Foshee, V. A., & Matthew, R. A. (2007). Adolescent Dating Abuse Perpetration: A Review of Findings, Methodological Limitations, and Suggestions for Future Research. In D. J. Flannery, A. T. Vazonsyi, & I. D. Waldman (Eds.), *The Cambridge Handbook of Violent Behavior and Aggression* (pp. 431-449). New York: Cambridge University Press

Foshee, V. A., & Reyes, H. L. M. (2009). Primary Prevention of Adolescent Dating Abuse: When to Begin, Whom to Target, and How to do it. In J. Lutzker & D. Whitaker (Eds.), *Preventing Partner Violence* (pp. 141-168): American Psychological Association

Foshee, V. A., Reyes, L. M., Agnew-Brune, C. B., Simon, T. R., Vagi, K. J., Lee, R. D., & Suchindran, C. (2014). The Effects of the Evidence-Based Safe Dates Dating Abuse Prevention Program on Other Youth Violence Outcomes. *Prevention Science: The Official Journal Of The Society For Prevention Research*. doi: 10.1007/s11121-014-0472-4

Fraker, T., & Maynard, R. (1987). The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs. *Journal of Human Resources, 22*(2), 194-227.

Franklin v. Gwinnett County (GA) Public Schools, 503 U.S. 60 (Supreme Court 1992).

Froot, K. A. (1989). Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. *Journal of Financial and Quantitative Analysis, 24*.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Gomez, A. M. (2010). Testing the Cycle of Violence Hypothesis: Child Abuse and Adolescent Dating Violence as Predictors of Intimate Partner Violence in Young Adulthood. *Youth & Society, 43*(1), 171-192. doi: 10.1177/0044118x09358313

Gomez, A. M. (2011). Testing the Cycle of Violence Hypothesis: Child Abuse and Adolescent Dating Violence as Predictors of Intimate Partner Violence in Young Adulthood. *Youth & Society, 43*(1), 171-192. doi: 10.1177/0044118x09358313

Gruber, J. E., & Fineran, S. (2008). Comparing the impact of bullying and sexual harassment victimization on the mental and physical health of adolescents. *Sex Roles, 59*(1-2), 1-13.

Hale, J. L., Householder, B. J., & Greene, K. L. (2003). The theory of reasoned action. In J. P. Dillard & M. Pfau (Eds.), *The persuasion handbook: Developments in theory and practice* (pp. 259-286). Thousand Oaks, CA: Sage

Halpern, C. T., Spriggs, A. L., Martin, S. L., & Kupper, L. L. (2009). Patterns of Intimate Partner Violence Victimization from Adolescence to Young Adulthood in a Nationally Representative Sample. *Journal of Adolescent Health, 45*(5), 508-516. doi: 10.1016/j.jadohealth.2009.03.011

Hamby, S., Finkelhor, D., & Turner, H. (2012). Teen Dating Violence: Co-Occurrence With Other Victimizations in the National Survey of Children's Exposure to Violence (NatSCEV). *Psychology of Violence, 2*(2), 111-124. doi: 10.1037/a0027191

Hand, J. Z., & Sanchez, L. (2000). Badgering or bantering? Gender differences in experience of, and reactions to, sexual harassment among US high school students. *Gender & Society, 14*(6), 718-746.

Hanson, M. J. (2010). Health Behavior in Adolescent Women Reporting and not Reporting Intimate Partner Violence. *Jognn-Journal of Obstetric Gynecologic and Neonatal Nursing, 39*(3), 263-276. doi: 10.1111/j.1552-6909.2010.01138.x

Hickman, L. J., Jaycox, L. H., & Aranoff, J. (2004). Dating Violence Among Adolescents: Prevalence, Gender Distribution, and Prevention Program Effectiveness. *Trauma, Violence, and Abuse, 5*, 123-142.

Hill, C., & Kearl, H. (2011). Crossing the line: sexual harassment in schools. Washington, DC: AAUW.

Hilton, N. Z., Harris, G. T., Rice, M. E., Krans, T. S., & Lavigne, S. E. (1998). Antiviolence education in high schools: Implementation and evaluation. *Journal of Interpersonal Violence, 13*, 726- 742.

Holt, M. K., & Espelage, D. L. (2005). Social support as a moderator between dating violence victimization and depression/anxiety among African American and Caucasian adolescents. *School Psychology Review, 34*(3), 309-328.

Howard, D. E., & Wang, M. Q. (2003a). Psychosocial factors associated with adolescent boys' reports of dating violence. *Adolescence, 38*(151), 519-533.

Howard, D. E., & Wang, M. Q. (2003b). Risk Profiles of Adolescent Girls who were Victims of Dating Violence. *Adolescence, 38*(149), 1-14.

Howard, D. E., Wang, M. Q., & Yan, F. (2007a). Prevalence and psychosocial correlates of forced sexual intercourse among US high school adolescents. *Adolescence, 42*(168), 629-643.

Howard, D. E., Wang, M. Q., & Yan, F. (2007b). Psychosocial factors associated with reports of physical dating violence among US adolescent females. *Adolescence, 42*(166), 311-324.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions.* Paper presented at the Fifth Berkeley Symposium on Mathematical Statistics and Probability.

Jackson, S. M., Cram, F., & Seymour, F. (2000). Violence and sexual coercion in high school students' dating relationships. *Journal of Family Violence, 15*(1), 23−36.

Jaffe, P. G., Sudermann, M., Reitzel, D., & Killips, S. M. (1992). An Evaluation of a Secondary School Primary Prevention Program on Violence in Relationships. *Violence and Victims, 7*, 129-146.

Jaycox, L. H., McCaffrey, D., Eiseman, B., Aronoff, J., Shelley, G. A., Collins, R. L., & Marshall, G. N. (2006). Impact of a school-based dating violence prevention program among Latino teens: Randomized controlled effectiveness trial. *Journal of Adolescent Health, 39*(5), 694-704.

Johnson, W. L., Giordano, P. C., Longmore, M. A., & Manning, W. D. (2014). Intimate Partner Violence and Depressive Symptoms during Adolescence and Young Adulthood. *Journal of Health and Social Behavior, 55*(1), 39-55. doi: 10.1177/0022146513520430

Jones, L. E. (1991). The Minnesota School Curriculum Project: A statewide domestic violence prevention project in secondary schools. In B. Levy (Ed.), *Dating violence: Young women in danger* (pp. 258−266). Seattle, WA: Seal Press

Jouriles, E. N., Garrido, E., Rosenfield, D., & McDonald, R. (2009). Experiences of psychological and physical aggression in adolescent romantic relationships: Links to psychological distress. *Child Abuse & Neglect, 33*(7), 451-460. doi: http://dx.doi.org/10.1016/j.chiabu.2008.11.005

Jouriles, E. N., Mueller, V., Rosenfield, D., McDonald, R., & Dodson, M. C. (2012). Teens' experiences of harsh parenting and exposure to severe intimate partner violence: Adding insult to injury in predicting teen dating violence. *Psychology of Violence, 2*(2), 125.

Jouriles, E. N., Platt, C., & McDonald, R. (2009). Violence in Adolescent Dating Relationships. *The Prevention Researcher, 16*, 3-7.

Krajewski, S. S., Rybarik, M. F., Dosch, M. F., & Gilmore, G. D. (1996). Results of a curriculum intervention with seventh graders regarding violence in relationships. *Journal of Family Violence, 11*(2), 93-112. doi: 10.1007/bf02336664

Kreiter, S. R., Krowchuk, D. P., Woods, C. R., Sinal, S. H., Lawless, M. R., & DuRant, R. H. (1999). Gender differences in risk behaviors among adolescents who experience date fighting. *Pediatrics, 104*(6), 1286-1292.

Lalonde, R. J. (1986). Evaluating the Econometric Evaluations of Training-Programs with Experimental-Data. *American Economic Review, 76*(4), 604-620.

Larkin, J. (1994). Walking through walls: The sexual harassment of high school girls. *Gender and Education, 6*(3), 263-280.

Lavoie, F., Vezina, L., Piche, C., & Boivin, M. (1995). Evaluation of a Prevention Program for Violence in Teen Dating Relationships. *Journal of Interpersonal Violence, 10*(4), 516-524.

Lee, V. E., Croninger, R. G., Linn, E., & Chen, X. (1996). The culture of sexual harassment in secondary schools. *American Educational Research Journal, 33*(2), 383-417.

Lipsey, M. W., & Wilson, D. B. (1993). The Efficacy of Psychological, Educational, and Behavioral Treatment - Confirmation from Metaanalysis. *American Psychologist, 48*(12), 1181-1209.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Lisa H. Jaycox, Daniel F. McCaffrey, Beverly Weidmer Ocampo, Gene A. Shelley, Susan M. Blake, Donna J. Peterson, . . . Kub, J. E. (2006). Challenges in the evaluation and implementation of school-based prevention and intervention programs on sensitive topics. *American Journal of Evaluation, 27*(3), 320-336.

MacGowan, M. J. (1997). An Evaluation of a Dating Violence Prevention Program for Middle School Students. *Violence and Victimization, 12*, 223-235.

Malik, S., Sorenson, S. B., & Aneshensel, C. S. (1997). Community and dating violence among adolescents: Perpetration and victimization. *Journal of Adolescent Health, 21*(5), 291-302.

Manganello, J. A. (2008). Teens, dating violence, and media use - A review of the literature and conceptual model for future research. *Trauma Violence & Abuse, 9*(1), 3-18. doi: 10.1177/1524838007309804

McDonald, P., Graham, T., & Martin, B. (2010). Outrage management in cases of sexual harassment as revealed in judicial decisions. *Psychology of Women Quarterly, 34*(2), 165-180.

McGahee, W., Kemp, V., & Tingen, M. (2000). A Theoretical Model for Smoking Prevention Studies in Preteen Children. *Prediatric Nursing, 26*, 135-143.

McMahon, S., & Banyard, V. L. (2011). When can I help? A conceptual framework for the prevention of sexual violence through bystander intervention. *Trauma, Violence, & Abuse*, 1524838011426015.

McMaster, L. E., Connolly, J., Pepler, D., & Craig, W. M. (2002). Peer to peer sexual harassment in early adolescence: A developmental perspective. *Development and Psychopathology, 14*(1), 91-105.

Mendelson, T., Turner, A. K., & Tandon, S. D. (2010). Violence exposure and depressive symptoms among adolescents and young adults disconnected from school and work. *Journal of Community Psychology, 38*(5), 607-621. doi: 10.1002/jcop.20384

Meyer, H., & Stein, N. (2004). Relationship Violence Prevention Education in Schools: What's Working, What's Getting in the Way, and What are Some Future Directions. *American Journal of Health Education, 35*(4), 198-205.

Miller, E., Tancredi, D. J., McCauley, H. L., Decker, M. R., Virata, M. C. D., Anderson, H. A., . . . Silverman, J. G. (2013). One-Year Follow-Up of a Coach-Delivered Dating Violence Prevention Program A Cluster Randomized Controlled Trial. *American Journal of Preventive Medicine, 45*(1), 108-112. doi: 10.1016/j.amepre.2013.03.007

Miller, E., Tancredi, D. J., McCauley, H. L., Decker, M. R., Virata, M. C. D., Anderson, H. A., . . . Silverman, J. G. (2013). One-year follow-up of a coach-delivered dating violence prevention program: a cluster randomized controlled trial. *American Journal of Preventive Medicine, 45*(1), 108-112.

Miller, E., Tancredi, D. J., McCauley, H. L., Decker, M. R., Virata, M. C. D., Anderson, H. A., . . . Silverman, J. G. (2012). "Coaching boys into men": a cluster-randomized controlled trial of a dating violence prevention program. *Journal of Adolescent Health, 51*(5), 431-438.

Molidor, C., & Tolman, R. M. (1998). Gender and contextual factors in adolescent dating violence. *Violence against Women, 4*(2), 180-194.

Mulford, C., & Giordano, P. C. (2008a). Teen Dating Violence: A Closer Look at Adolescent Romantic Relationships. Washington, DC: National Institute of Justice.

Mulford, C., & Giordano, P. C. (2008b). *Teen Dating Violence: A Closer Look at Adolescent Romantic Relationships* (Vol. Issue No. 261). Washington, DC: National Institute of Justice.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Nagayama Hall, G. C., & Barongan, C. (1997). Prevention of Sexual Aggression: Sociocultural Risk and Protective Factors. *American Psychological Association, 52*(1), 5-14.

Nahapetyan, L., Orpinas, P., Song, X., & Holland, K. (2014). Longitudinal Association of Suicidal Ideation and Physical Dating Violence among High School Students. *Journal of Youth and Adolescence, 43*(4), 629-640. doi: 10.1007/s10964-013-0006-6

National Institute of Justice (Producer). (2011, 3/28/13). Teen Dating Violence.

Noonan, R. K., & Charles, D. (2009a). Developing Teen Dating Violence Prevention Strategies Formative Research With Middle School Youth. *Violence against Women, 15*(9), 1087-1105. doi: 10.1177/1077801209340761

Noonan, R. K., & Charles, D. (2009b). Developing Teen Dating Violence Prevention Strategies: Formative Research With Middle School Youth. *Violence against Women, 15*(9), 1087-1105. doi: 10.1177/1077801209340761

O'Callaghan, F. V., Chant, D. C., Callan, V. J., & Baglioni, A. (1997). Models of alcohol use by young adults: An examination of various attitude-behavior theories. *Journal of Studies on Alcohol, 58*(5), 502-507.

O'Keefe, M. (1997). Predictors of Dating Violence Among High School Students. *Journal of Interpersonal Violence, 12*, 546-568.

O'Leary, K. D., & Slep, A. M. S. (2003). A dyadic longitudinal model of adolescent dating aggression. *Journal of Clinical Child and Adolescent Psychology, 32*(3), 314-327. doi: 10.1207/S15374424JCCP3203_01

Offenhauer, P., & Buchalter, A. (2011). Teen Dating Violence: A Literature Review and Annotated Bibliography: Report Prepared by the Federal Research Division, Library of Congress under an Interagency Agreement with the Violence and Victimization Research Division, National Institute of Justice.

Pacifici, C., Stoolmiller, M., & Nelson, C. (2001). Evaluating a prevention program for teenagers on sexual coercion: A differential effectiveness approach. *Journal of Consulting and Clinical Psychology, 69*, 552- 559.

Patton, M. Q. (1990). *Qualitative Evaluation and Research Methods*. Newbury Park: Sage.

Pellegrini, A. D. (2001). A longitudinal study of heterosexual relationships, aggression, and sexual harassment during the transition from primary school through middle school. *Journal of Applied Developmental Psychology, 22*(2), 119-133.

Pepler, D. J., Craig, W. M., Connolly, J. A., Yuile, A., McMaster, L., & Jiang, D. (2006). A developmental perspective on bullying. *Aggressive Behavior, 32*(4), 376-384. doi: 10.1002/ab.20136

Peskin, M. F., Markham, C. M., Shegog, R., Baumler, E. R., Addy, R. C., & Tortolero, S. R. (2014). Effects of the It's Your Game . . . Keep It Real Program on Dating Violence in Ethnic-Minority Middle School Youths: A Group Randomized Trial. *American Journal of Public Health, 104*(8), 1471-1477. doi: 10.2105/AJPH.2014.301902

Potter, S. J., & Moynihan, M. M. (2011). Bringing in the bystander in-person prevention program to a us military installation: results from a pilot study. *Military medicine, 176*(8), 870-875.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology, 27*(1), 85-96.

Riecken, H. W., Boruch, R. F., Campbell, D. T., Caplan, N., Glennan, T. K., Pratt, J. W., . . . Williams, W. (1974). *Social Experimentation: A Method for Planning and Evaluating Social Programs*. New York: Academic Press.

Roberts, T. A., & Klein, J. (2003). Intimate partner abuse and high-risk behavior in adolescents. *Archives of Pediatrics & Adolescent Medicine, 157*(4), 375-380.

Roberts, T. A., Klein, J. D., & Fisher, S. (2003). Longitudinal effect of intimate partner abuse on high-risk behavior among adolescents. *Archives of Pediatrics & Adolescent Medicine, 157*(9), 875-881.

Rogers, W. H. (1993). Regression standard errors in clustered samples. (Vol. 3, pp. 88-94): Stata Technical Bulletin

Schewe, P. A. (2000a). Evaluating the effects of ICASA's 29 existing rape prevention programs. Retrieved June 22, 2011, from http://tigger.uic.edu/~schewepa/web-content/newpages/evaluations.html

Schewe, P. A. (2000b). *Report of Results of the STAR Project to the Illinois Violence Prevention Authority*. Chicago.

Schewe, P. A. (Ed.). (2002). *Guidelines for developing rape prevention and risk reduction interventions: Lessons from evaluation research*. Washington, D.C.: American Psychological Association.

Sears, H. A., Byers, E. S., & Price, E. L. (2007). The Co-Occurrence of Adolescent Boys' and Girls' use of Psychological, Physically, and Sexually Abusive Behaviors in their Dating Relationships. *Journal of Adolescence, 30*, 487-504.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Desiigns for Generalized Causal Inference*. New York: Houghton Mifflin Company.

Shanklin, S. L., Brener, N., McManus, T., Kinchen, S., & Kann, L. (2007). *2005 Middle School Youth Risk Behavior Survey*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Sherman, L. W. (1992). *Policing Domestic Violence: Experiments and Dilemmas*. New York: Free Press.

Shrier, L. A., Pierce, J. D., Emans, S. J., & DuRant, R. H. (1998). Gender differences in risk behaviors associated with forced or pressured sex. *Archives of Pediatrics & Adolescent Medicine, 152*(1), 57-63.

Silverman, J. G., Raj, A., & Clements, K. (2004). Dating violence and associated sexual risk and pregnancy among adolescent girls in the United States. *Pediatrics, 114*(2), 220-225.

Silverman, J. G., Raj, A., Mucci, L. A., & Hathaway, J. E. (2001). Dating violence against adolescent girls and associated substance use, unhealthy weight control, sexual risk behavior, pregnancy, and suicidality. *JAMA-Journal of the American Medical Association, 286*(5), 572-579.

Silverschanz, P., Cortina, L. M., Konik, J., & Magley, V. J. (2008). Slurs, snubs, and queer jokes: Incidence and impact of heterosexist harassment in academia. *Sex Roles, 58*(3-4), 179-191.

Smith, P. H., White, J. W., & Moracco, K. E. (2009). Becoming who we are: a theoretical explanation of gendered social structures and social networks that shape adolescent interpersonal aggression. *Psychology of Women Quarterly, 33*(1), 25-29. doi: 10.1111/j.1471-6402.2008.01470.x

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Spoth, R. L., Redmond, C., & Shin, C. (2001). Randomized trial of brief family interventions for general populations: adolescent substance use outcomes 4 years following baseline. *Journal of Consulting and Clinical Psychology, 69*(4), 627.

Spriggs, A. L., Halpern, C. T., & Martin, S. L. (2009). Continuity of adolescent and early adult partner violence victimisation: association with witnessing violent crime in adolescence. *Journal of Epidemiology and Community Health, 63*(9), 741-748. doi: 10.1136/jech.2008.078592

Stein, N. (1981). Sexual harassment of high school students: Preliminary research results. Boston, MA: Massachusetts Department of Education.

Stein, N. (1995). Sexual Harassment in K-12 Schools: The Public Performance of Gendered Violence. *The Harvard Educational Review, 65*(2), 145-162.

Stein, N. (1999). *Classrooms and Courtrooms: Facing Sexual Harassment in K-12 Schools*. New York: Teacher's College Press.

Stein, N., Marshall, N., & Tropp, L. (1993). Secrets in public: Sexual harassment in our schools. A report on the results of a Seventeen magazine survey. Wellesley, MA: Wellesley College Center for Research on Women.

Straus, S. (1988). Sexual harassment in the school: Legal implications for principals (Vol. 72 (506), pp. 93-97). Reston, VA: National Association of Secondary School Principals.

Taylor, B., Mumford, E., Liu, W., & Stein, N. (2015). Assessing different levels and dosages of the Shifting Boundaries intervention to prevent youth dating violence in New York City middle schools: A randomized control trial. Washington DC: National Institute of Justice.

Taylor, B., & Stein, N. (Ohio 2005-2007). Dating Violence Prevention Programs in NYC Middle Schools: A Multi-Level Experiment: National Institute of Justice.

Taylor, B., Stein, N., & Burden, F. (2010a). The Effects of Gender Violence/Harrassment Prevention Programming in Middle Schools: A Randomized Experimental Evaluation. *Violence and Victims, 25*, 202-223.

Taylor, B., Stein, N., Mack, A. R., Horwood, T. J., & Burden, F. (2008). *Experimental Evaluation of Gender Violence/Harassment Prevention Programs in Middle Schools*. Washington DC: National Institute of Justice.

Taylor, B., Stein, N., Mumford, E., & Woods, D. (2013a). Shifting Boundaries: An Experimental Evaluation of a Dating Violence Prevention Program in Middle Schools. *Prevention Science, 14*(1), 64-76. doi: 10.1007/s11121-012-0293-2

Taylor, B., Stein, N. D., Woods, D., & Mumford, E. (2011). Shifting Boundaries: Final Report on an Experimental Evaluation of a Youth Dating Violence Prevention Program in New York City Middle Schools. Washington, DC: National Institute of Justice.

Taylor, B., Stein, N., & Burden, F. (2010). The Effects of Gender Violence/Harrassment Prevention Programming in Middle Schools: A Randomized Experimental Evaluation. *Violence and Victims, 25*, 202-223.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Taylor, B. G., & Mumford, E. A. (2014 e-pub ahead of print). A national descriptive portrait of adolescent relationship abuse: Results from the National Survey on Teen Relationships and Intimate Violence (STRiV). *Journal of Interpersonal Violence*, 1-26. doi: 10.1177/0886260514564070

Taylor, B. G., Mumford, E. A., Hair, E. C., Stein, N., & Yu, T.-C. (2010). *Dating Violence Prevention Programs in New York City Public Middle Schools: A Multi-level Experimental Evaluation*. Abstract submitted to the 2011 Annual Conference of the Society for Research in Child Development.

Taylor, B. G., Stein, N., & Burden, F. (2010b). Exploring gender differences in dating violence/harassment prevention programming in middle schools: Results from a randomized experiment. *Journal of Experimental Criminology, 6*(4), 419–445. doi: 10.1007/s11292-010-9103-7

Taylor, B. G., Stein, N. D., Mumford, E. A., & Woods, D. (2013b). Shifting Boundaries: An experimental evaluation of a dating violence prevention program in middle schools. *Prevention Science, 14*(1), 64-76. doi: 10.1007/s11121-012-0293-2

Teitelman, A. M., Ratcliffe, S. J., Dichter, M. E., & Sullivan, C. M. (2008). Recent and past intimate partner abuse and HIV risk among young women. *Journal of Obstetric, Gynecologic, & Neonatal Nursing, 37*(2), 219-227.

Tolman, D. L., Spencer, R., Rosen-Reynoso, M., & Porche, M. V. (2003). Sowing the seeds of violence in heterosexual relationships: Early adolescents narrate compulsory heterosexuality. *Journal of Social Issues, 59*(1), 159-178.

U.S. Department of Education, Office for Civil Rights (2001, January 19). Revised sexual harassment guidance: Harassment of students by school employees, other students, or third parties. From the Office of the Assistant Secretary of Civil Rights. Available from: http://www2.ed.gov/about/offices/list/ocr/docs/shguide.html.

US Department of Education, Office for Civil Rights. March 13, 1997. Sexual harassment guidance: Harassment of students by school employees, other students or third parties. Notice 62 Federal Register 12034-12051.

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research, 16*(3), 219-242.

Van Dulmen, M. H., Klipfel, K. M., Mata, A. D., Schinka, K. C., Claxton, S. E., Swahn, M. H., & Bossarte, R. M. (2012). Cross-lagged effects between intimate partner violence victimization and suicidality from adolescence into adulthood. *Journal of Adolescent Health, 51*(5), 510-516.

Ward, K. J. (2002). *Making the invisible visible: A feminist evaluation of an adolescent gender violence prevention program.* Doctoral dissertation. Cornell University. Ithaca, NY.

Weisburd, D., Petrosino, A., & Mason, G. (1991). When bigger is not better: Design sensitivity in a sample of criminal justice experiments. In N. Morris & M. Tonry (Eds.), *Crime and justice: An annual review of research* (Vol. 15). Chicago: The University of Chicago Press

Weisz, A. N., & Black, B. M. (2001). Evaluating a sexual assault and dating violence prevention program for urban youth. *Social Work Research, 25*(2), 89- 100.

Wekerle, C., & Wolfe, D. A. (1999). Dating violence in mid-adolescence: Theory, significance, and emerging prevention initiatives. *Clinical Psychology Review, 19*(4), 435-456.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Whitaker, D. J., Morrison, S., Lindquist, C., Hawkins, S. R., O'Neil, J. A., Nesius, A. M., . . . Reese, L. (2006). A critical review of interventions for the primary prevention of perpetration of partner violence. *Aggression and Violent Behavior, 11*(2), 151-166. doi: 10.1016/j.avb.2005.07.007

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*, 817-830.

White House Task Force to Protect Students from Sexual Assault. (2014). Not Alone. Washington, D.C.: White House Task Force to Protect Students from Sexual Assault.

Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics, 56*, 645-646.

Wolfe, D. A., Crooks, C., Jaffe, P., Chiodo, D., Hughes, R., Ellis, W., . . . Donner, A. (2009). A School-Based Program to Prevent Adolescent Dating Violence A Cluster Randomized Trial. *Archives of Pediatrics & Adolescent Medicine, 163*(8), 692-699. doi: 10.1001/archpediatrics.2009.69

Wolfe, D. A., Scott, K., Reitzel-Jaffe, D., Wekerle, C., Grasley, C., & Straatman, A. L. (2001). Development and validation of the conflict in adolescent dating relationships inventory. *Psychological Assessment, 13*(2), 277-293. doi: 10.1037//1040-3590.13.2.277

Wolfe, D. A., Wekerle, C., Scott, K., Straatman, A. L., Grasley, C., & Reitzel-Jaffe, D. (2003). Dating violence prevention with at-risk youth: A controlled outcome evaluation. *Journal of Consulting and Clinical Psychology, 71*(2), 279-291. doi: 10.1037/0022-006x.71.2.279

Wolitzky-Taylor, K. B., Ruggiero, K. J., Danielson, C. K., Resnick, H. S., Hanson, R. F., Smith, D. W., . . . Kilpatrick, D. G. (2008a). Prevalence and correlates of dating violence in a national sample of adolescents. *Journal of the American Academy of child and Adolescent Psychiatry, 47*(7), 755-762. doi: 10.1097/CHI.0b013e318172ef5f

Wolitzky-Taylor, K. B., Ruggiero, M., Danielson, C. K., Resnick, H. S., Hanson, R. F., Smith, D. W., . . . Kilpatrick, D. G. (2008b). Prevalence and correlates of dating violence in a national sample of adolescents. *Journal of the American Academy of child and Adolescent Psychiatry, 47*(7), 755-762. doi: 10.1097/CHI.0b013e318172ef5f

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Appendices

Appendix 1: Shifting Boundaries Intervention for NYC-2

Appendix 2: Survey instruments used for the Shifting Boundaries NYC-2 evaluation

This document is a research report submitted to the U.S. Department of Justice. This report has not been published by the Department. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.