Georgia State University
Digital Archive @ GSU

Psychology Theses | Department of Psychology

5-27-2008

Fidelity-Outcomes Relationships in the Expect Respect Program

Angela Devi Mooss
amooss@hotmail.com

Follow this and additional works at: http://digitalarchive.gsu.edu/psych_theses

Recommended Citation
Mooss, Angela Devi, "Fidelity-Outcomes Relationships in the Expect Respect Program" (2008). *Psychology Theses.* Paper 49.

This Thesis is brought to you for free and open access by the Department of Psychology at Digital Archive @ GSU. It has been accepted for inclusion in Psychology Theses by an authorized administrator of Digital Archive @ GSU. For more information, please contact digitalarchive@gsu.edu.

FIDELITY-OUTCOMES RELATIONSHIPS IN THE EXPECT RESPECT PROGRAM

by

ANGELA D. MOOSS

Under the Direction of James G. Emshoff, Ph.D.

ABSTRACT

The effects of program fidelity, gender, socioeconomic status, and school level were tested on various outcomes of a dating violence prevention program, Expect Respect. Fidelity data was collected from program facilitators, and individual posttest scores were gathered for individuals nested within each programmatic site. Multiple HLM models indicated that main effects for fidelity were present for the knowledge gained outcome scale, such that higher program fidelity led to higher posttest scores for participants. No other site level predictors affected outcomes or the fidelity-outcome relationship. Results from this study point towards the importance of implementing program fidelity when cognitive gains are a central goal of the program, whereas a more flexible program approach may be more optimal in conveying other programmatic components.

INDEX WORDS: Fidelity, Implementation, Prevention, Dating violence, Expect Respect, Hierarchical Linear Modeling

FIDELITY-OUTCOMES RELATIONSHIPS IN THE EXPECT RESPECT PROGRAM

by

ANGELA D. MOOSS

A Thesis submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2008

Copyright by
Angela Devi Mooss
2008

FIDELITY-OUTCOMES RELATIONSHIPS IN THE EXPECT RESPECT PROGRAM

by

ANGELA D. MOOSS

Committee Chair: James G. Emshoff

Committee: Gabriel Kuperminc
Julia Perilla

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2008

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES vii

CHAPTER

1 INTRODUCTION 1

Program Implementation 2

Program Fidelity 3

Program Adaptation 4

Fidelity Assessments 6

Measuring Fidelity 8

Fidelity Criteria and Core Components 10

Fidelity, Adaptation, and Outcomes 11

Assessing Fidelity in Dating Violence Prevention Programs 14

Expect Respect 15

2 RESEARCH HYPOTHESES 17

Theory-Driven 17

Exploratory/Bi-directional 17

3 METHOD 18

Participants 18

Measures 19

Procedures 21

4 RESULTS 24

Data Analysis Strategy 24

Descriptive Data 25

Missing Data 26

Computing Variables 26

ANCOVA Models 27

Abusive Behaviors 28

Knowledge Gained 29

5 DISCUSSION 35

Limitations 38

Conclusions and Future Directions 39

REFERENCES 42

APPENDICES 47

A FACILITATOR FIDELITY INTERVIEW 47

B PRE-POST STUDENT SURVEY 57

LIST OF TABLES

Table 1: Expect Respect student and facilitator demographics and frequencies by site location 23

Table 2: Descriptive data on variables of interest 32

Table 3: HLM analyses of the effects of moderators on the fidelity-outcomes relationship 33

LIST OF FIGURES

Figure 1: The effects of fidelity only on the knowledge gained outcome 34

Fidelity-Outcomes Relationships in the Expect Respect Program

Introduction

Consistent with social and health sciences literature is the notion that the amount of time it takes to transfer successful research into practice is excessive and potentially harmful. Numerous interventions designed for community improvements have been developed and evaluated. However, not only is there a gap between what research says is effective and what is developed programmatically, there is also a gap between the efficacy a program demonstrates, and its implementation into the wider community (Wandersman, 2003; Glasgow & Emmons, 2007). A program designed to prevent dating violence, for example may have successful outcomes in the context in which it was developed, but these effects may not transfer if the program is adopted by a new organization. These situations result in the inefficient use of limited resources (Graham, Logan, Harrison, Straus, Tetroe, Caswell, & Robinson, 2006). Participant and organizational differences in new program settings both contribute to different, sometimes less successful programmatic outcomes. Aspects of the program may not be culturally relevant to new populations and organizations may lack some of the resources necessary to implement the new program exactly as it was designed. Disseminating effective interventions into different settings is a complex process in which the program must be adopted, implemented, and sustained in a relevant manner to produce the same positive effects it did in the original setting. Discovering which aspects of the program, organizations, and participants maximize positive outcomes can be a first step towards educating leaders on which programs they should choose to adopt and how they should be implemented.

Policy makers have indicated an increased need for evidence-based programs to be implemented for prevention, and social sciences research supports this assertion. Several reports

issued by the Institute of Medicine (IOM) discussed problems with the current levels of health care people were receiving and sought to fill the gap between the care people could receive and what they were actually getting (2001). To do this, the IOM emphasized incorporating research results into practices, thus creating the idea of evidence based practices (EBP's). EBP's mean that a practice is based on a substantial body of evidence (usually 2 or more efficacy studies, plus establishment of effectiveness in a diverse setting). Additionally, evidence-based practices must include tools to assist in implementation and to monitor quality and performance. Such tools facilitate fidelity to model's critical ingredients and increase the likelihood of producing outcomes found in the efficacy studies (Bond, Becker, & Drake, 2001; Mowbray, Bybee, Holter, & Lewandowski, 2006). The tools provided to organizations that choose to adopt a program may include training manuals, curriculum, videos, evaluation measurements, and some form of technical assistance. Once the program is adopted, however, numerous questions arise as to how the program should be implemented and how program staff should be trained (Rohrbach, Dent, Skara, Sun, & Sussman, In Press).

Program Implementation

Implementation is the "delivery, uptake, and context of the intervention (Panzano & Roth, 2006). Implementation follows adoption of a program and includes all aspects of how the program is run in a new setting. Implementation has been called a critical gateway between adoption and the routine use of an innovation (Klein and Sorra, 1996). Understanding the process of implementation is useful for psychologists to better understand specifics of a program that made it successful in its original context. Unfortunately, implementation is rarely a straightforward process. Many interventions that have displayed successful outcomes were not consistently tested throughout the implementation process; therefore, little is known about which

programmatic factors can be modified without divorcing the program from its essence. There is a need to provide descriptive information about programs in evaluation studies, but such information is rarely included (Bond, et. al, 2001).

Organizational resources, participant demographics and characteristics, training and staff experience, and inter-organizational relationships can all influence how a program is implemented, and therefore need to be described in sufficient detail to organizations choosing to adopt a program. Additionally, such information about the implementation of a program is crucial for replication of such programs, but process evaluations are rarely as praised as outcome evaluations, especially when program stakeholders are heavily invested and time is an issue. Finally, the collection of such data is rarely planned for at the time of the program's inception, meaning the program may not be amenable to such comprehensive data collection (Brekke & Wolkon, 1998). The implementation process is complex and comprehensive, including all aspects of a program from adoption through delivery. The purpose of this research is to focus on program fidelity. Fidelity is defined as the degree to which program providers implement programs as intended by program developers (Dusenbury, Brannigan, Falco, Hansen, 2003).

Program Fidelity

Program fidelity, implementing a program exactly as it is intended by developers, can be conceptualized in different ways according to the goals of the researcher. For example, fidelity related to treatment design is intended to test the degree to which the hypotheses can adequately be tested in relation to the underlying theory. Additionally, fidelity related to program receipt tests whether participants receiving services have understood that they have received services and are able to perform skills obtained as intended. Finally, fidelity related to delivery serves to establish whether the core structures and processes of the intervention were delivered as intended

by the program model (Resnick, Bellg, & Borelli, 2005). The focus of this study is on both program delivery and program receipt. Delivery and receipt of treatments are generally the most common types of fidelity that are evaluated; however, over half of the studies appearing in literature (psychology, behavior therapy, marital and family, and psychiatry) ignored issues of treatment fidelity (Summerfelt, 2003). Furthermore, Peterson, Homer, and Wonderlich found only 20% of studies discussed fidelity (1982); and Rogers-Weise found that only 6% of studies actually measured fidelity (1992).

Implementation is a complicated process that requires support and buy-in of organizations and other stakeholders. It is important for organizations and implementers to stay on track and recognize common implementation problems in a timely and effective manner. During initial stages of implementation, success is often associated with fidelity to the evidence-based practice or program. After the program has been implemented, fidelity can also contribute to the understanding of outcomes. For example, ineffective programs can be implemented well and effective programs can be implemented poorly, both leading to negative outcomes. Truly desirable outcomes can only be achieved when programs are both effective and well-implemented, indicating the importance of both program evaluations and fidelity assessments.

Program Adaptation

Programs that are presumed to be effective may not produce positive results once disseminated to different settings. In many circumstances, institutions such as schools that are most in need of interventions are under-resourced. When prevention programs are adopted by less affluent sites, they may need to be modified to fit with the organization's resources. Another perspective on implementation, program adaptation, occurs any time the adopting organization

alters the way the program is run based on the needs of the population or the resources of the organization.

Changes/adaptations may be reflected on the individual, organizational, and community levels. For example, at an individual level, individual participants may have a different need for the program (i.e. secondary prevention versus primary prevention); therefore, it might be necessary to alter the exclusion/inclusion criteria for participants. Organizational level factors may include tailoring the program to a size the current staff can handle. For example, some schools may lack the capacity to hire outside program facilitators, and may have to deliver the program when their schedule allows. Finally, programmatic components often need to be modified to insure that original program successes are generalizeable to alternative settings. Such changes may entail eliminating parts of the program or modifying certain aspects that may be culturally inappropriate or unnecessary. Unfortunately, when communities and adopting sites are viewed as passive, their assets and knowledge on their own needs is bypassed in favor of a program model that ignores the value of indigenous practices (Miller and Shinn, 2005).

Adaptations happen for a number of reasons, and can be both intended and accidental. Often, there is a mismatch between what scientists design and what communities have the capacity to implement (Miller and Shinn, 2005). Programs are designed with ideals in mind about staff experience and availability, organizational resources, and participant willingness. Unfortunately, organizations that have the greatest need to implement such programs generally have the fewest resources. Even when adopting organizations lack the resources necessary to implement exactly as intended, modifications to programs are usually minor and do not affect the program's core components. Research has shown that over half of the organizations that adopt

programs generally make adaptations, although only about 20% of these changes are wide-scale (Dusenbury, et al., 2003).

Although special population needs may be a reason that some researchers do not require implementers to strive for perfect fidelity, tailoring programs to fit with communities is not widely accepted. Evidence-based programs are presumed to have benefits over indigenous practices because most of the latter have not been extensively evaluated (Miller and Shinn, 2005). These views describe what Rogers defined as the pro-innovation bias (1995). This bias leads to an inaccurate view of whether and how effectiveness is evaluated in community and organizational change processes. Miller and Shinn suggest that the IOM's (2001) model of dissemination is too simplistic and that much could be learned from alternative approaches, such as learning what works from communities, and recognizing that external, not internal, validity is the crux of program evaluation because it allows one to address the relevance of an innovation (Cronbach, 1980).

Fidelity Assessments

Fidelity assessments have a number of important purposes, but are mainly used to measure how effectively a program has been implemented. Assessments can be used to describe the degree to which prescribed program elements were included and non-prescribed elements were absent from implementation. Also, fidelity measurements can serve as evidence to policy-makers that programs are being implemented as intended, and resources are being utilized by the target audience (Mowbray, Holter, Teague, & Bybee, 2003). Without fidelity assessments, when program outcomes fall short, there is no way to establish whether this was due to the inadequacy of the program model, or a lack of skilled implementation. Furthermore, an understanding of

fidelity can lead to the dissemination of higher quality implementation practices (Dusenbury, Brannigan, et al., 2003).

Fidelity sheds light on how feasible a program is for particular settings. If the program model requires that nearly all components be implemented for effectiveness, certain organizations may choose not to adopt based on resources and population need. To fully capture the implementation process, fidelity assessments need to be comprehensive. Ideally, an assessment of fidelity should be able to identify whether program elements leading to successful outcomes were included, whether there were elements included that were not intended by the original model, and the degree to which interventions that differ on salience service parameters need to be assessed (Summerfelt, 2003).

Fidelity can also be looked at as a way to better operationally define the independent variable. For example, the effect of a dating violence prevention program may be intended to produce increases in participant's knowledge of abuse. The dependent variable is explicated through scores on a defined measure, but the program, the independent variable, is rarely subjected to a manipulation check. Manipulation checks involve testing the varying levels of the independent variable to insure that each level is categorically different from the others. Manipulation checks are used to show that an independent variable has construct validity, and are particularly useful when interpreting what has happened when a study does not produce the predicted results on the dependent variable. Additionally, without a manipulation check, the researcher cannot identify whether the program failed to have an effect, or whether it had an effect, just not the one predicted (Hoyle, Harris, & Judd, 2001).

Using fidelity assessments as manipulation checks can also serve as a caution against committing type III errors. A type III error occurs when "…faulty measurements, experimental

designs, or conceptualizations of crucial variables prohibit meaningful interpretation of experimental results" (Smith & Sechrest, 1991, p. 237). Type III errors are most closely associated with construct validity, and can also occur when variables of interest are not adequately defined or poorly reported. Researchers pay a great deal of attention to operationalizing and defining the dependent variable, but often neglect to adequately describe the nature and process of the independent variable and its presumed effects (Summerfelt, 2003). Because fidelity assessments can describe in detail how a program/treatment works, they can be used to combat type III errors.

Well-developed and valid fidelity measures can improve statistical power in treatment outcomes, acting as moderator variables to help explain variance in outcomes (Mowbray, et al., 2003). Furthermore, problems with understanding how a program works are especially problematic when implementers attempt to disseminate findings from an effective program (Smith & Sechrest, 1991). Therefore, type III errors also endanger the external validity of a study if programmatic results lead to flaws in generalizability. Finally, fidelity assessments can be used to improve external validity by outlining the procedures and guidelines for replicating the model, such as characteristics of the participants and setting (Mowbray, et al., 2005). Because fidelity assessments can serve to improve both internal validity through operationalizing the independent variable and external validity by specifying replication guidelines, measuring fidelity is essential to truly evaluate a program's effects.

Measuring Fidelity

Measuring fidelity often occurs both during and after a program has run its course. When fidelity to an original model is measured early on in the implementation process, it can serve as a checkpoint to program implementers about how well they are replicating the major components

of a program. This information can be useful both for making any changes to current program practice to boost fidelity, and for documenting instances in which fidelity is either not possible or not a goal of implementation. Process notes such as these can serve other organizations looking to adopt with similar setting differences or resource constraints.

The most common two methods to measure fidelity are 1) ratings by experts based on project documentation, client records, site observations, and video-taped sessions and 2) surveys or interviews completed by those delivering the services or receiving them (Mowbray, et al., 2003). Examples of fidelity data collection directly from participants are rarer and more methodologically diverse. Unrau (2001) suggests that fidelity data from clients should not be the only type of fidelity information gathered, but rather should complement other sources of data. Consistent with this approach, not all items can be validated by participants because clients will not know about all of the activities that are intended and occur throughout a program.

Calsyn (2000) outlined methods that could be used to assess the validity of fidelity assessments. Most fidelity measures have content validity because they were developed by experts and forerunners in the field. Other studies have cited that these measures also have predictive validity based on the positive outcome of participants. Unfortunately, this method is extremely problematic for a number of reasons. As mentioned earlier, past research on fidelity and program outcomes is not consistent; indicating that for some populations, fidelity to a program model could have detrimental effects. Additionally, the notion that high fidelity predicts better outcomes eliminates the potential to study variables that moderate the relationship between fidelity and outcome. Regardless of which method is used, the core components of a program (fidelity criteria) need to be identified and operationalized prior to developing an adequate measure.

Fidelity Criteria and Core Components

Three major steps in establishing fidelity criteria have been defined (McGrew, et al., 1994). The first is to identify critical components of a model through expert consensus and developer input. The second is to collect data to measure those components, and the third step is to then examine the reliability and validity by assessing data collected. The empirical approach to determine what elements of a program must be included (cannot be adapted) to maintain positive outcomes, involves deconstructing all core components of a program and testing the impact of each component on programmatic outcomes (Kelly, Heckman, Stevenson, & Williams, 2000).

Calsyn (2000) notes that a single, summed fidelity score is not an accurate measure of services delivered in a program because two programs could receive the same score, but be very different in how they operated on site; therefore, literature suggests that fidelity criteria should include aspects of both structure and process. Structure encompasses the framework for service delivery and process components reflect the way in which services are delivered. Each of these process and structure elements can further be divided into subcomponents that can be measured by a more detailed fidelity assessment. For example, Borelli and David (2004) recognized training implementers as one of the sections that should be included on a fidelity instrument for health behavior research. Implementer use of the same materials and curriculum as developers, in addition to ongoing communication with program developers are also dimensions that may be included on fidelity measures. Ultimately, however, fidelity instruments are designed to be program specific and should measure the core components of an intervention.

Gresham (1993) has argued that fidelity can actually be quantified by calculating the percentage of treatment core components implemented, but it is often difficult to determine

which program components are essential without conducting an intense review (McGrew, Bond, Dietzen, & Salyers, 1994). Additionally, quantifying fidelity, reduces the complex implementation process to statistical analyses, and ignores the addition of any elements (Eyesenck, 1984). Cohesive with this idea, many researchers also suggest that rather than measure fidelity formally, reports should explain the implementation of innovations in sufficient detail for readers to replicate and apply their results. This view further emphasizes the idea that documentation of any modifications is necessary for comprehensive evaluation of the program and its model.

Upon adopting a program, it becomes crucial for implementers to understand what the essential and non essential-components of a program are so that they can maintain adherence while adapting the program to fit their local need. Allowing staff to make adaptations as long as they do not compromise the underling theory of the program is a more flexible approach that implementers can take. The most effective strategy for achieving this is for the program developers and site to consistently communicate with each other to insure that essential components are not absent and that any modifications and additions are secondary to the required components. When this cooperation occurs, program fidelity can be measured according to not only the core components that were implemented, but also the adaptations that were made (including both enhancements and deletions) giving a detailed summary of the implementation process. This information can then be used to insure that replication of the most effective program can be achieved.

Fidelity, Adaptation, and Outcomes

Research about the importance of maintaining high levels of fidelity in implementation is mixed. Traditionally high fidelity has been conceptualized as a gold standard because it is rooted

in evidence and efficacy trials. A common finding in psychotherapy literature is that interventions adhering more closely to prescribed program standards are more effective (Bond, et al., 1997). Prevention programs have assessed fidelity only slightly more than psychotherapy interventions have, although results are often similar, indicating that high fidelity leads to better participant outcomes. For example, Dusenbury (2003) found that the more completely a provider implements a program, the more successful the outcome will be, and that when programs are not implemented as intended, they are less likely to be effective. Furthermore, research has indicated that when key elements are left out of implementation, negative and contradictory outcomes can result (Bond, et al., 1997).

Specific evidence linking high fidelity to better student outcomes in drug abuse prevention programs has also been noted (Dusenbury, et al., 2004). Additionally, a study by Battistich, Schaps, Watson, and Solomon found that higher fidelity predicted greater improvements in children's behavior and attitudes towards drug use, whereas lower fidelity did not show these gains. In fact, these programmatic outcomes were not even recognized by researchers until sites were divided into low versus high fidelity, citing the importance of including fidelity assessments in all program evaluations (2000).

Although the majority of past research dictates a need for high fidelity in program implementation, lower fidelity has also been linked to successful outcomes, particularly when cultural sensitivity is an issue (Wandersman, 2003). A more recent emphasis on the specific needs of participants has raised questions about whether too much fidelity has the potential to harm a population. In one study, high fidelity implementation was associated with more positive outcomes, but also that local additions to the program increased effectiveness, suggesting that as

long as main program components are there, additions do not decrease effectiveness (Blakely, Mayer, Gottschalk, Schmitt, Davidson, Roitman, & Emshoff, 1987).

Conclusively linking high fidelity with better programmatic outcomes is difficult because whether or not perfect fidelity is the goal for program implementers, adaptations almost always occur. Szulanski and Winter (2002) say that adapting a successful program template is a mistake, but evidence that a program works does not mean that the program will have a smaller effect if it is only partially implemented, implemented differently, or that it will cause harm either way. Additionally, there may be a significant need to tailor programs to fit cultural and community needs (Hohmann and Shear, 2002). The question then becomes: Are program modifications errors as most researchers/evaluators believe, or are they intended, necessary, and appropriate to fit the local need? (Ridgely & Jerrell, 1996).

Program modifications are often made during the implementation process in response to needs that may not have been recognized at the start of the program or from unpromising outcome data collected early on. This flexible approach to program implementation reflects some components of Participatory Action Research, and relies on the implementer to continuously alter the "path" of the program to optimize participant outcomes. Unfortunately, this complicates outcome implications because it may be unclear whether success is due to participant comfort and willingness to follow the program versus actual elements of the program model (Summerfelt, 2003). Furthermore, if a program strays too far from the model intervention and core components are eliminated or altered, outcome data should be excluded from any meta-analysis of program effect (Mowbray, et al., 2003).

Assessing Fidelity in Dating Violence Prevention Programs

The context of the current study focuses on the fidelity of a dating violence prevention program implemented throughout middle schools and high schools in Texas and Ohio. Although dating violence and sexual assault continue to exist at alarming rates in society, few programs have been developed to address the issue; and even fewer have been evaluated. Even after such prevention programs are evaluated, a gap still exists between what has proven effective and what is actually done at a programmatic setting. Adolescents need to be educated about sexual assault and dating violence using programs that have been successfully evaluated; however, the lack of empirically supported knowledge about school and contextual conditions that support implementation and sustainability is striking (Shinn, 2003).

Sexual assault and dating violence is a pervasive topic in American culture today. For younger adolescents, the numbers of people who have experienced dating violence are astounding, and college experience estimates continue to increase at alarming rates. Although much needed research has been done on college populations' experiences of sexual assault and dating violence, far less has been conducted with younger students in middle school and high school (Weisz & Black, 2001). Research indicates that 57% of high school students reported engaging in at least one aggressive act against a dating partner within the past year (Avery-leaf, Cascardi, O'leary, 2002). Not only does this suggest a dating violence epidemic, the numbers are also troubling because adolescents are less likely to seek out help from service providers.

Because school settings are an optimal way to reach large numbers of youth at once, organizations have begun to develop prevention programs that address the need for sexual assault and dating violence education and prevention. The creation of these new programs

necessitates that they be evaluated thoroughly before being disseminated to other school settings and environments.

Research on the effectiveness of such programs in raising student knowledge and awareness about dating violence provides mixed results. Avery-leaf, et al. found significant decreases in pro-violence attitudes among all participants at a European-American high school, but failed to follow up to determine potential long-term effects (2002). Safe dates, a school-based dating violence prevention program, documented significant reductions in participants' violent behaviors and changes in cognitive risk factors. At one year follow up, cognitive changes were still evident; however other positive outcomes had waned (Foshee, Bauman, Ennett, Linder, Benefield, & Suchindran. 2004). These results are consistent with those from other adolescent prevention evaluations in that behavioral effects fade, but changes in cognitive risk factors are maintained (Foshee, Bauman, Greene, Koch, Linder, & MacDougall, 2000). Finally, although some studies report differences in program outcomes according to gender (Jaffe, Suderman, & Reitzel, 1992), the Safe Dates program was equally effective for males and females and for whites and non-whites (Foshee, et al., 2000). The variation in findings and lack of consistent follow- up necessitate the need for program implementers and evaluators to carefully document the implementation process with regards to fidelity and program adaptations. Without fidelity assessments, when unsuccessful outcomes occur, it is difficult to tell whether the model itself failed to have long-term effects, or whether staff implementing the program made adaptations that lessened its effects.

Expect Respect

Expect Respect, a dating violence prevention program has been successfully evaluated; however, little research has been conducted about the fidelity of the implementation of the

Expect Respect curriculum. The program works towards supporting youth who have experienced past abuse, raising youth expectations for respect in relationships, and promoting youth leadership in violence prevention. Specifically, Expect Respect implementers are trained to model healthy behavior and serve as mentors in an environment where student participants can learn and practice healthy relationship skills (Ball, Kerig, & Rosenbluth, in press). The program consists of a curriculum, materials (discussion questions and activities) and evaluation materials that cover various topics related to healthy relationships over the course of 18-24 sessions. Groups are generally 6-10 individuals and are separated by gender. An evaluation funded by the Centers for Disease Control and Prevention measured change in students' knowledge, attitudes, beliefs, and behaviors related to healthy relationships (Safeplace, 2006). Results of the evaluation indicated that positive changes occurred in students' ability to identify abusive behaviors. Specifically, knowledge of abusive and controlling behavior increased for both boys' and girls' groups; however, slight differences in the types of behaviors they identified were noted (Safeplace, 2006). Unfortunately, although gender differences in similar prevention program outcomes have been studied, results remain too few to be conclusive. Furthermore, specific outcomes in sexual assault prevention programs have rarely been studied in relation to their implementation. The fidelity with which implementers ran such programs could reveal knowledge about why particular outcomes occur and may shed light on why gender and other differences in such outcomes are present.

This study will examine how fidelity, gender, socio-economic status, school level, and outcomes all relate to one another regarding a gender-based dating violence prevention intervention. Specifically, this study seeks to explore how fidelity can serve as a predictor for posttest scores of participants in the Expect Respect program. Additionally, gender, school level,

and SES will be tested as moderators between the fidelity – outcome relationship to determine whether any interactions exist. Because Expect Respect is a gender-based intervention that also separates middles school and high school groups, it is logical to conclude that facilitators may have implemented the program differently for girl versus boy groups or middle schools versus high school groups to hold participant interest and involvement. Furthermore, the socio-economic status of the school environment may contribute to the fidelity-outcome relationship in terms of resources available, and familiarity to the participants. Because low socioeconomic status is a risk factor for youth violence (Centers for Disease Control and Prevention, 2007), facilitators may need to alter their implementation strategies depending upon which school they are in. The various outcomes for program participants that will be tested include coping, abusive behaviors, healthy behaviors, and knowledge gained.

Research Hypotheses

Theory-driven

- The fidelity with which Expect Respect facilitators implemented the program will predict student participants' posttest scores. Specifically, higher fidelity will be linked to better outcome scores (when controlling for pretest scores).

Exploratory/Bidirectional

- The gender of the Expect Respect group will serve as a moderator between the fidelity – outcome relationship.
- The school level of the Expect Respect group will serve as a moderator between the fidelity-outcome relationship.
- The socio-economic status of the Expect Respect site will serve as a moderator between the fidelity-outcome relationship.

Method

Participants

Participants for this study included both Expect Respect facilitators and student participants. Program facilitator contact information was obtained through Safeplace Organization and reflected programmatic sites in Austin, Texas; Toledo, Ohio; and Oxford, Ohio. SafePlace is an organization that provides community outreach, education, and prevention services to help individuals and families better identify and address sexual violence issues. Group facilitators were recruited primarily through Safeplace Organizations in Austin, Texas, and Miami University in Oxford, Ohio. The Lucas County Domestic Violence Prevention Enhancement and Leadership through Alliances project (DELTA) recruited in Toledo, Ohio.

Expect Respect facilitators (N = 30) were mostly female (70%) and conducting groups in Austin. The majority of facilitators identified as Caucasian (63.7%), and other racial/ethnic groups represented included Black/African-American (13%), Latino/Hispanic (13%), and Multiracial (10%). Additionally, nearly all group leaders (90%) had experience working with teen groups, and all participants had experience leading small groups. Furthermore, Safeplace organization provided facilitators full materials including: the Expect Respect training manual, curriculum for 18 - 24 sessions, materials to be used in activities, discussion questions for all sessions, and evaluation materials. Training sessions were held for all group facilitators by program developers, and technical assistance was also available throughout the implementation process. Finally, schools in which Expect Respect programs were implemented were those that Safeplace organization, Miami University, and the DELTA project had relationships with, either through conducting other school-based interventions, or providing community support. For final

analyses, three facilitators were not used (*N* = 27), because not enough individual pretest and posttest score data matched up with those particular groups.

At the individual level, respondents included those students involved in an Expect Respect program (*N* = 168). These student participants had been identified by school faculty and/or staff members as having experienced, witnessed, perpetrated or were at risk of experiencing domestic, sexual, or dating violence. Of these students, 60% were in high school (as opposed to middle school) and 67% were female. Racial/ethnic data was also collected on students in Expect Respect groups. Participants most frequently identified as Caucasian (39.3%), Hispanic/Latino (29.7%), and Black/African American (19%). Other racial/ethnic categories endorsed included Native Hawaiian/Pacific Islander, American Indian/Alaskan Native, Multiracial, and other. Participants in the Expect Respect program attended an average of 15 sessions. Demographics and other data regarding participants and group facilitators can be found in Table 1.

Furthermore, an average of 50% of students across schools was eligible for the National School Lunch Program (NSLP), although percentages ranged from 7% to 90%. The proportion of students per school who qualified for the NSLP was used as the socio-economic status variable for a particular Expect Respect group, as has been done in other studies (Centers for Disease Control and Prevention, 2007). These data were obtained for each school throughout Austin and Ohio from The United States Department of Agriculture Food and Nutrition Services website (2007).

Measures

Facilitator Fidelity Interview

Instruments to assess fidelity were developed based on both extensive conversations with program developers identifying core components of the program, and the criteria by which their degree of implementation could be judged. Fidelity instruments were created with Expect Respect program developers through phone and in-person interviews after an extensive review of the training manual. Researchers also observed facilitator training sessions offered by Safeplace. The interviews contained both open and close-ended questions that assessed a variety of fidelity dimensions including; facilitator training, program processes, participation in sessions, group process, and technical assistance (See Appendix A). The Cronbach Alpha reliability estimate was relatively low for the fidelity measure (α = .65). This was expected given that the items represented a count of the number of core components that were implemented with fidelity to the original model, thus higher fidelity to one component does not preclude high fidelity to others. Additionally, although the Program Fidelity Interview captured the various core component dimensions of the Expect Respect program, for the purposes of this study, only a single mean score for fidelity was used.

Fidelity between the original program and the manner in which it was implemented by SafePlace was measured on a High, Medium, Low scale, whereby "High" fidelity was given a score of +1, "Medium" fidelity a score of 0, and "Low" fidelity a score of – 1. On average, each interview lasted about 45 minutes. Hall and Loucks (1978) developed this method for measuring fidelity and it has been cited in numerous other fidelity research studies (Blakely, Emshoff, and Roitman, 1984; Blakely, et al., 1987; Emshoff, Blakely, & Gray, 2003).

Pre-Post Student Outcome Survey

Safeplace Organization developed a pretest/ posttest survey to administer to student program participants during the first and last sessions of the program. Questions on the survey

were taken from a number of various instruments and scales that measure issues around teenage relationships and dating including: conflict, perceptions, rejection and vulnerability, interpersonal difficulties, problem-solving, support-seeking, and coping (Safeplace, 2006). Specifically, questions asked participants whether they viewed certain behaviors as controlling or abusive, as well as asked students to identify what they did during an after an argument/fight with a boyfriend or girlfriend (See Appendix B). The current study used survey items that corresponded to the following scales: knowledge of healthy behaviors (n=8), knowledge of abusive behaviors (n=12), knowledge gained (n=9), and coping strategies (n=5). These particular outcome scales were chosen based on the logic model outcomes developed by Safeplace (Ball, Kerig & Rosenbluth, in press). The Cronbach Alpha reliability estimates for pre- and post-test were .78 and .68 for knowledge of healthy behaviors scale, .89 and .85 for the knowledge of abusive behaviors scale, and .90 and .88 for knowledge gained scale. The Cronbach Alpha reliability estimate was low on the coping strategies scale (pre-test: α= .59 and post-test: α= .57) which is expected given that the items on this measure assessed a variety of different coping methods, thus the endorsement of one coping strategy did not necessitate the use of another. Additionally, pretest and posttest variables had small ranges with "1" as the minimum and "4" as the maximum for the abusive and healthy scales or "5" as the maximum for the coping and knowledge scales.

Procedures

Expect Respect group facilitators administered the Expect Respect pretest survey to student participants during the first session of the program, and students completed the posttest survey during the final session. Each survey took approximately 20 minutes to complete. Upon completion of posttests, students were matched to pretests via name, and were then coded by

program facilitator to maintain confidentiality. Pre-post Student Survey data was sent to Safeplace for initial data entry and analysis before being sent to Georgia State University.

Phone interviews were conducted with program facilitator participants at each site over the course of two years regarding the degree of fidelity to which they implemented their Expect Respect groups. Facilitators who ran multiple groups were often contacted on two occasions to discuss their fidelity in each group separately. Expect Respect groups generally started at the onset of the school year (September) and at the beginning of the calendar year (January); therefore, facilitator participants were mostly contacted for interviews towards the end of the program, either in November or April. All interviews were conducted over the telephone by trained graduate students who assisted in the development of the interviews. Informed consent was obtained from all participants at the beginning of each interview and participant responses to interview questions were recorded as they occurred.

Table 1

Expect Respect student and facilitator demographics and frequencies by site location.

	Oxford, Ohio		Toledo, Ohio		Austin, Texas	
	Freq (N)	Percent	Freq (N)	Percent	Freq (N)	Percent
Student Participants						
Male	19	35.2	7	70.0	29	27.9
High School	37	68.5	10	100.0	54	51.9
Caucasian	40	74.1	1	10.0	25	24.0
Hispanic/Latino	0	0	0	0	50	48.1
Black/African-American	13	24.1	7	70.0	12	11.5
Native Haw./Pacific Isl.	0	0	0	0	1	1.0
Amer. Ind./Alaska Nat.	0	0	0	0	1	1.0
Multiracial	1	1.9	0	0	9	8.7
Other	0	0	2	20.0	1	1.0
Missing	0	0	0	0	5	4.8
Total	54	100	10	100	104	100

	Oxford, Ohio		Toledo, Ohio		Austin, Texas	
	Freq (N)	Percent	Freq (N)	Percent	Freq (N)	Percent
Facilitator Participants						
Male	2	22.2	2	40.0	5	31.3
Caucasian	9	100	4	80.0	6	37.5
Hispanic/Latino	0	0	0	0	4	25.0
Black/African-American	0	0	1	20.0	3	18.8
Multiracial	0	0	0	0	3	18.8
Experience with teen groups	7	77.8	5	100.0	15	93.8
Experience with either teens or groups	2	22.2	0	0	1	6.3
Total	9	100	5	100	16	100

Results

Data Analysis Strategy

Many analyses that aim to compare the fidelity of a program (at a group level) with individual outcome scores have ignored within group variability. Similarly, analysis at the individual level ignores the fact that individual units are not independent of similar program/group effects. Neither of these approaches to analyzing nested data is optimal or appropriate; therefore, current statistical software has allowed for a more accurate, comprehensive approach to studying variance in multilevel, nested data (Mowbray, et al., 2003).

Multilevel modeling techniques (Hierarchical Linear Modeling) allow the researcher to incorporate substantive theory about organizational effects at different levels into the clustered nature of individual survey data (Heck & Thomas, 2000). Specifically, the researcher can examine links between variables measured at an individual level and variables measured at the organizational level in which individuals are nested. In this study, the outcome variables of coping, abusive behaviors, healthy behaviors, and knowledge gained were all measured on the individual level. Program fidelity, gender, school level, and socio-economic status are all group level variables because there is no within group variability for any of them. For example, within a particular group, gender is consistent, as is grade level, and SES is consistent because it was measured at the school level. Additionally, because only one facilitator per group was interviewed, participants within a group were all matched to the same fidelity score.

The data collected was analyzed utilizing the HLM software (Bryk, Raudenbush, & Condon, 1996) to test hypotheses across individual and organizational levels. The HLM model included Expect Respect participants' pretest scores as a level 1 covariate and posttest scores as a level 1 outcome variable (there were no individual level 1 predictors). Next, gender, fidelity, and

the gender*fidelity interaction term were entered into the model as level two predictors of the level 1 outcome. Two other identical models were also tested using either school level or socio-economic status instead of gender and their specific interaction effects with fidelity. These two level intercepts-as-outcomes model were conducted to estimate cross level effects of fidelity, gender, school level, and socio-economic status on program-related outcomes. Specifically, analyses examined the intercepts of a particular outcome at the individual level (coping, abusive behaviors, healthy behaviors, knowledge gained) as a function of group level variables, controlling for pretest scores on that outcome. Analyses were run using maximum marginal likelihood estimation for mixed-effects regression models using HLM 6 software (Raudenbush, Bryk, & Condon, 1996).

Descriptive Data

As mentioned previously, the sample of student participants included slightly more females than males and more students in high school than middle school sites. Additionally, the proportion of students at program sites that qualified for a federally subsidized school lunch was used as a setting level marker of economic disadvantage. By this measure, sites varied in socioeconomic status, with 7% to 90% of students economically disadvantaged. The final site level predictor, fidelity of Expect Respect groups, varied between .36 and .85 (1 represented perfect fidelity), and the average fidelity score was .62 (*SD*. = .13). Furthermore, posttest scores ranged from 2.89 (healthy behaviors) to 4.12 (knowledge gained) depending on their ranges (See Table 2). Finally, all variables tested in the HLM models showed normal distributions with normal residual distributions.

Missing Data

HLM software is only able to handle missing data from level-one data sources, which it can address by eliminating missing variables listwise and pairwise (Heck & Thomas, 2000). Within the level-one SPSS database, a large portion of data was incomplete (38%). Most commonly missing variables included either a pre or posttest score from a student participant. These individuals who only completed one or the other were manually deleted at the beginning of the data-cleaning process. Furthermore, individuals who failed to complete demographic data were also eliminated from the initial database. An additional problem was created by individuals who failed to answer all questions on the pretest or posttest. Cases where more than 20% of data was missing from pre-posttest answers were deleted (n = 3). After these cases were deleted, 165 cases remained. Data was imputed (less than 20%) to the remaining cases prior to creating the HLM level-one database, in order to maximize the use and accuracy of available data. To establish whether the values were missing randomly, Little's MCAR test was conducted in SPSS. Results indicated that data were missing completely at random, X^2 (12) = 11.473, p = .489. Missing values were therefore imputed using the expectation maximization algorithm (EM). Because facilitator participant interviews were administered by trained graduate students, any missing level two data was obtained with follow up emails.

Computing Variables

Prior to analyzing the data with HLM software, certain variables were combined and centered to create a readable database. First, all continuous variables were centered around their mean for ease in interpreting results. These variables included "fidelity" and "SES". Additionally, to test the interaction effects of fidelity and gender, fidelity and middle/high school, and fidelity and SES, interaction variables were created. These separate variables were

calculated by multiplying the fidelity score (post-centering) by the appropriate moderator. For this study, gender and school level were dummy coded so that for gender, “0” represented female and “1” represented male, and for school level, “0” represented middle school and “1” represented high school.

ANCOVA models

Prior to testing any models using HLM, it was necessary to determine whether adequate between group and within group variance in the variables of interest existed. A series of analyses of covariance (ANCOVAs) were computed in order to estimate whether there was significant between- and within-group variance in each outcome variable. In each model, the dependent variable was a program outcome, the independent variable was group membership, and the covariate was the pre-test score for the outcome variable. By partitioning all of the variance in outcomes into between group and within group components, the ANCOVA model enables the researcher to calculate an intra-class correlation (Heck and Thomas, 2000). Chi square significance tests with $p < .05$ imply there is sufficient variance between groups to be accounted for by a level two variable. The covariates (pre-test scores for each outcome) were grand-mean centered upon being entered into the model. This means that each score was centered around the mean for all participants, as opposed to the mean of participants within a particular group. HLM software completes this step during the analysis.

Consistent with HLM, ANCOVA models were run for all four dependent variables before testing any level two predictors. Results of the ANCOVA model revealed that two of the outcome variables lacked sufficient between-group variance to test level two predictors. The healthy behaviors intra-class correlation was not significantly different from zero, indicating that only a negligible amount of the variability in the healthy behaviors outcome can be attributed to

a level two predictor, X^2 (26) = 23.93, $p > .500$. Additionally, the coping outcome intra-class correlation was also not significantly different from zero, X^2 (26) = 36.36, $p = .09$. The other two outcomes, abusive behaviors and knowledge gained, demonstrated high enough intra-class correlations to continue running HLM analyses using the level 2 predictors of fidelity, gender, school level, and SES. First, the abusive behaviors scale had an intra-class correlation of .197, X^2 (26) = 65.64, $p < .001$. Finally, amount of knowledge gained by Expect Respect participants also had a significant intra-class correlation, .15, X^2 (26) = 53.57, $p = .001$.

Abusive behaviors

After the ANCOVA model revealed enough between group variance to test level two predictors on abusive behaviors, three separate models were tested to establish whether a relationship between fidelity and outcomes was moderated by another variable. These models are known as "intercept as outcome" models because they test the moderation effects on the intercept of the dependent variable (Bryk, Raudenbush, & Condon, 1996). In this study, there is no level one predictor (individual level predictor); therefore, testing a moderation effect on the slope of the level one predictor-outcome is not possible.

First, the level one covariate (pretest) and outcome are entered in level one variables. Next, the level two variables (fidelity, a moderator, and interaction term) are entered at the intercept of the outcome. Because the continuous variables (fidelity and SES) were centered prior to entering them in HLM, and dichotomous variables (gender and middle/high school) can be interpreted at the point of zero, all variables were entered "uncentered". Below is an example of the HLM model as described where Y is the outcome variable, B0 is the level one intercept of the outcome within the level two unit, B1 is the level one covariate, and R is the error.

Level-1 Model

Y = B0 + B1* (ABUSIVE) + R

Level-2 Model

B0 = G00 + G01* (GENDER) + G02* (FIDELITY) + G03* (GEN*FID) +U0

B1 = G10

The intercept as outcome model for all three moderators revealed no significant results. Main effects for fidelity, $p = .119$, and gender, $p = .974$, were both non-significant. The fidelity*gender interaction was not significant either, T (23, 7) = 1.568, $p = .130$. No significant effects were found when SES was added into the model. Main effects for fidelity, $p = .322$, SES, $p = .205$, and the fidelity*SES interaction, T (23, 7) = -.008, $p = .994$, revealed no significant effects. Finally, when the school level variable was tested as a moderator on abusive behavior outcomes, no significant effects were found for the main effects of fidelity, $p = .345$, school level, $p = .173$, or the interaction, T (23, 7) = .620, $p = .541$. A complete listing of B values and standard errors for all results can be found in Table 3.

Knowledge gained

To test whether a significant moderation effect existed between the fidelity-outcome relationship regarding knowledge gained during the Expect Respect program, three HLM models were tested. These intercepts as outcome models tested both the main effects and interaction effects of fidelity and either gender, SES, or school level. No significant main effects were obtained for school level, $p = .125$, and the school level and fidelity interaction was also not significant, T (23, 7) = .584, $p = .558$.

The intercepts as outcome model that tested the effects of gender and fidelity revealed a main effect for fidelity, T (23, 7) = 2.057, $p = .026$. With each unit increase in fidelity scores,

posttest scores for the knowledge gained scale increased by .96 points. Neither the main effect for gender, $p = .201$, nor the interaction, $p = .666$, were significant.

Additionally, when the main effects of fidelity and SES were run, the fidelity main effect was significant again, $T(23, 7) = 3.138$, $p = .003$. As fidelity increased by one unit, scores for the knowledge gained posttest increased by .97 points. The main effect for SES, $p = .202$ and the interaction term, $p = .372$, were both non-significant (See Table 3).

HLM intercepts as outcome models are most concerned with a model's fit to the data, rather than a change in variance accounted for, or R^2. Finding a main effect for fidelity in two of three models, when knowledge gained was the dependent variable, provides evidence that a better model fit may just include fidelity, and not other level two predictors or interaction terms. To test this additional *a priori* hypothesis, two further analyses were conducted. First, an HLM model was created predicting only fidelity as a level two predictor of knowledge gained. Next, a Chi square deviance test for model fit was conducted (Bryk & Raudenbush, 2002). This test establishes whether the deviance from the model as stated is reduced significantly when fewer variables are predicted and the model is simpler (i.e. fidelity only versus fidelity, gender, and fidelity * gender).

Results of the fidelity-only model revealed a significant effect of fidelity on knowledge gained. However, a trend in the main effect of fidelity did exist, $T(25, 5) = 1.884$, $p = .036$ (one-tailed). This trend suggests that those groups that operated with above average fidelity (with respect to other Expect Respect groups) scored .73 points higher ($SE = .39$) on the posttest outcome of knowledge gained than those groups with lower fidelity (See Figure 1).

First, the Chi square deviance test was performed on the results of the gender-fidelity model. There were no differences in model fit between the simpler model testing only fidelity and the more complex model testing fidelity and gender, $X^2 = 2.36$, $p > .05$.

A chi square deviance test was also performed to establish whether a fidelity only model was more accurate than a model including fidelity, SES, and the fidelity-SES interaction. Results of this analysis once again revealed no significant differences existed between the two models, $X^2 = 2.51$, $p > .05$.

These deviance tests provide justification that a more parsimonious model including only fidelity as the predictor of knowledge gained is the best model for this data. Because no significant differences existed between the fidelity-only model and other models which included additional moderators, the simpler model is the most accurate (Bryk, Raudenbush, & Condon, 1996). Fidelity proved to be a significant predictor of individual outcomes when run alone, therefore, the aforementioned results support the first hypothesis. That is, fidelity predicts individual outcomes, such that higher fidelity leads to more positive outcomes.

Table 2

Descriptive data on variables of interest

Variable	N	Minimum	Maximum	Mean	SD
Level Two					
Fidelity	27	.36	.85	.62	.14
SES	27	.07	.90	.50	.27
Gender	27	0.0	1.00	.33	.48
School Level	27	0.0	1.00	.63	.49
Outcomes					
Posttest Abusive	165	1.00	4.00	3.35	.56
Posttest Healthy	165	1.00	4.00	2.89	.53
Posttest Coping	165	1.40	5.00	3.21	.73
Posttest Knowledge	165	1.11	5.00	4.12	.81

Table 3

HLM analyses of the effects of moderators on the fidelity-outcomes relationship

Outcome	Level two Predictor (s)	Coefficient	SE	*T*-ratio	*P*
Abusive behavior	Gender	-.004	.114	-.033	.974
	Fidelity	-.499	.510	-.978	.119
	Gender*fidelity	1.321	.842	1.568	.130
	School level	-.155	.110	-1.406	.173
	Fidelity	-.323	.576	-.403	.345
	School level*fidelity	.495	.799	.620	.541
	SES	.271	.207	1.306	.205
	Fidelity	-.205	.438	-.468	.322
	SES*fidelity	-.015	1.864	-.008	.994
Outcome	Level two Predictor (s)	Coefficient	SE	*T*-ratio	*P*
Knowledge gained	Gender	-.232	.176	-1.318	.201
	Fidelity	.973	.473	2.057	.026*
	Gender*fidelity	-.404	.926	-.437	.666
	School level	.235	.148	1.590	.125
	Fidelity	.347	.770	.450	.328
	School level*fidelity	.637	1.073	.594	.558
	SES	-.417	.317	-1.314	.202
	Fidelity	.960	.306	3.138	.003*
	SES*fidelity	-1.706	1.873	-.911	.372

* Fidelity main effect *p* values are listed for a one-tailed significance test

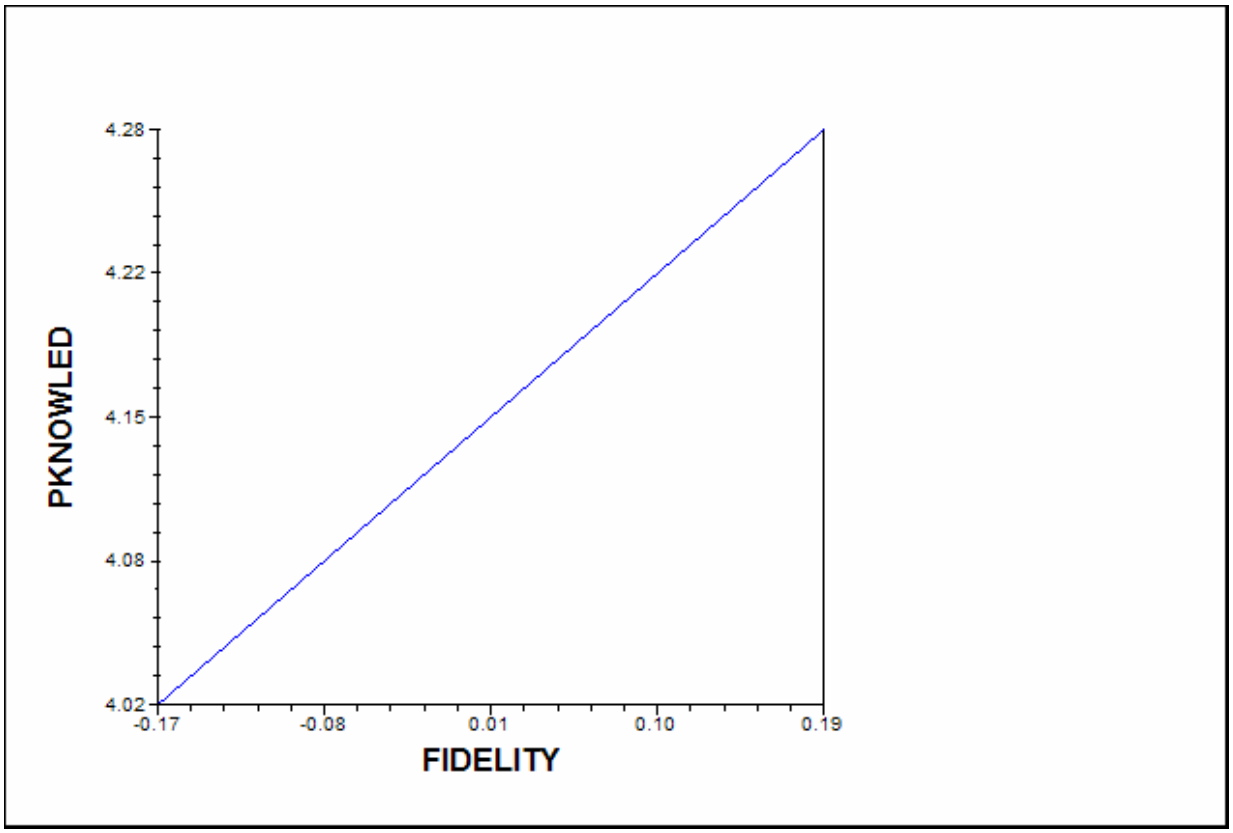

Figure 1. The effects of fidelity only on the knowledge gained outcome

Discussion

This study used HLM to examine the relationship between treatment fidelity (a site level variable) and four different individual level outcomes. Unexpectedly, two of the dependent variables, coping skills and healthy behaviors, had little variance in scores and could not be included in further analyses. This lack of variance could be due to the nature of the outcome scales. For example, a ceiling effect was evident in the healthy behaviors scale scores, indicating that these may already be recognized by most youth, and therefore little variance is seen in participant scores reflecting these topics. Unfortunately, because further analyses could not be conducted, additional research is needed to see if treatment fidelity affects these programmatic outcomes when adequate variance is present. Significant between group variance did exist in both the abusive behaviors and knowledge gained outcomes making it possible to examine the link between treatment fidelity and individual scores for these outcomes.

Posttest scores on the abusive behaviors outcome scales were not affected by fidelity or any site level variables (gender, school level, or SES). These results could be indicative of a number of things. First, although enough level one variance existed in abusive behavior posttest scores to warrant running more complex models, the range of this scale was the most limited, and had a relatively small standard deviation in comparison with other posttest scales. Furthermore, there was a very small point increase in mean scores between pretest and posttest. Because pretest scales were included as covariates in the HLM models, they tended to absorb most of the variance, leaving little to predict with site level variables. Furthermore, the individual data for the abusive behaviors scale shows evidence of a floor effect, indicating that that student participants were victims or perpetrators of these behaviors very little, thus little change was detected as a result of the program.

A significant main effect for fidelity was present on knowledge gained outcome scores, such that higher fidelity indicated greater amounts of cognitive gains. No other site level variables produced significant effects for this outcome. This finding is consistent with past research that demonstrated specific evidence linking high fidelity to better student outcomes in drug abuse prevention programs (Dusenbury, et al., 2004). Findings in this study are also consistent with much of the implementation literature that suggests higher fidelity leads to more promising outcomes (Dusenbury, Brannigan, et al., 2003; IOM, 2001; Bond, Becker, & Drake, 2001; Mowbray, Bybee, Holter, & Lewandowski, 2006). In general, although some researchers have expressed caution in regards to recommending high fidelity implementation (Miller & Shinn, 2005; Wandersman, 2003), most experts conclude that fidelity should be emphasized to program implementers, at least in regards to program theory and core components (Bond, et al., 1997; Blakely, et al., 1987; Mowbray, et al., 2003).

The impact of fidelity was only significant for the knowledge gained outcome, but did not affect abusive behavior scale scores. This finding suggests that there may be aspects of a program in which sticking to the curriculum and program design is more important. For example, implementers may need to use higher fidelity implementation when delivering a program that seeks to convey information and understanding about a specific topic, such as dating violence. Past research on dating violence prevention programs has found evidence in the importance of making cognitive gains among young participants, because often these outlast any behavior change outcomes (Foshee, Bauman, Ennett, Linder, Benefield, & Suchindran, 2004; Foshee, Bauman, Greene, Koch, Linder, & MacDougall, 2000).

Maintaining high fidelity regarding programmatic components meant to increase participant knowledge is crucial for long-term understanding about the dangers of dating

violence. Furthermore, the Theory of Reasoned Action asserts that knowledge changes are necessary for behavior change, and that changes in knowledge will eventually translate into attitude, intention, and eventually behavior change (Ajzen, 1991). Specifically, exposure to persuasive information causes progressive changes in knowledge, attitudes, and ultimately, behavior (Holtz & Thwombly, 2007). Consistent with this theory is the idea that a main goal of school-based prevention curriculum should be to build knowledge and awareness of an issue (dating violence) as a precursor to prevention of perpetration and victimization. Ajzen's Theory of Reasoned Action has been used as a basis for numerous prevention programs (Fishbein & Yzer, 2003), including sexual risk reduction (Albarracin, Johnson, Fishbein, & Muellerleile, 2001), drug abuse prevention (Holtz & Thwombly, 2007), and domestic violence prevention (Nabi, Southwell, & Hornik, 2002).

Gender, school level, and socioeconomic status did not affect participant scores for any Expect Respect programmatic outcomes. Although some studies report differences in program outcomes according to gender (Jaffe, Suderman, & Reitzel, 1992), the Safe Dates program was equally effective for males and females (Foshee, et al., 2000), and a past Expect Respect evaluation was also consistent with this study, finding no differences in gender specific to knowledge gained scale scores (Safeplace, 2006). This study's findings on school level are also consistent with past research on dating violence prevention programs (Center for Disease Control and Prevention, 2007), suggesting that the Expect Respect program is equally as effective for middle schools and high schools. Finally, although higher fidelity implementation was associated with better posttest scores for one programmatic outcome, gender, school level, and socioeconomic status had no effect on the fidelity-outcomes relationship. Thus, this study is

evidence that fidelity in implementing prevention programs is important for both boys and girls groups, as well as in all school environments, regardless of grade or socioeconomic factors.

Limitations

The results obtained in this study must be interpreted with caution as there were several limitations. First, although there was a relatively large N for level one individual posttest scores ($N = 165$), the number of sites at level two was low ($N = 27$). Calculating power for hierarchical models is generally very complex, and methods have only been developed for either longitudinal data or nested data involving randomly selected sites (Bryk & Raudenbush, 2002). Therefore, the minimum level two N was unspecified; however, larger numbers of participants indicate less opportunity for type I errors. Unfortunately, the small level two N in this study was also an issue because all of the predictors being tested were site level (level two) variables. Choosing level one predictors, variables that differed within Expect Respect groups, may have been a better option for formulating exploratory hypotheses for this data.

In addition to the small number of Expect Respect groups, there was also a lack of variability in outcomes being tested. Two of the four outcomes, coping and healthy behaviors, were unable to be run in the proposed model because of this issue. The lack of significant effects on the other dependent variables could have been caused by the amount of variance absorbed by pretest variables run as covariates. As stated before, the lack of change in participant scores could also be due to student's knowledge and experience prior to participating in the Expect Respect program, as both ceiling and floor effects were present in the data.

The main site level predictor, fidelity, also had little variance between groups. The overall mean for fidelity indicated that most group facilitators implemented the Expect Respect program according to developers' standards, although not stringently. This notion was consistent

with qualitative data collected from a larger Expect Respect study, in which facilitators shared that they understood the importance of following the curricula, but also felt the freedom to go off-track slightly in order to maintain high levels of participant satisfaction and interest. Facilitators reported being trained to stay true to the Expect Respect program messages, and that their goal was to relay those in any way they saw fit. Although Expect Respect facilitators had flexibility with regards to how they implemented the program, fidelity scores remained fairly similar, leaving little variability between sites. Fortunately, this freedom to run Expect Respect groups flexibly may have prevented a social desirability bias in data collected from facilitator interviews.

Finally, although in this study fidelity was measured as a single construct, it is a complex variable that has multiple parts. For example, site characteristics and implementer characteristics can also be aspects of fidelity, depending on how program developers define their core components. Implementer training, education, and ethnicity, as well as site readiness and site support are also aspects of fidelity that can affect program outcomes. Unfortunately, due to the nature of facilitator recruitment and program site selection, there was little variation in these aspects of fidelity for the Expect Respect program and implementers, thus further research about how the various dimensions of fidelity can influence particular outcomes may lend more insight as to what developers should stress as the most important implementation standards, as well as how much fidelity is necessary for the most optimal outcomes.

Conclusions and Future Directions

Conclusively linking high fidelity with better programmatic outcomes is difficult because whether or not perfect fidelity is the goal for program implementers, adaptations almost always occur. Although some developers train implementers to maintain a strict adherence to the

program curriculum (Elliott & Mihalic, 2004), others are emphasizing the importance of a flexible model that can be adapted to specific population needs (Ringwalt, Ennett, Vincus, & Simons-Rudolph, 2004). Understanding the aims of the program developers is an important factor in research surrounding program fidelity and outcomes. Results from this study indicate that the nature of programs seems to play a role in how developers should train implementers with regards to fidelity. Because knowledge gained by participants was linked to high fidelity, programs which seek to improve participant understanding of a certain topic and whose main components include standardized bodies of information to be imparted to participants need to be delivered with high fidelity for the most optimal outcomes. Other programs, such as those with behavior and action skills as main outcome goals may need to include a more malleable curriculum and program design, allowing facilitators and participants to reach an optimal method of program delivery.

Because evidence on how treatment fidelity can affect program outcomes is mixed, more studies examining the nature of the relationship between these two variables need to be conducted to determine when developers should stress fidelity to program implementers, versus when a more flexible implementing approach is most optimal. A flexible approach to program implementation is consistent with a Participatory Action Research model that allows for program modifications to occur during the implementation process in response to needs that may not have been recognized at the start of the program, but are made evident through repeated reflection and dialogue between the researcher and researched (Knightbridge, King, & Rolfe, 2006). Although some researchers recognize that flexible program designs may increase the adoption of the intervention (Glasgow & Emmons, 2007), fidelity to core aspects within a program may still be important. Thus, communication between program developers and implementers is crucial to

insure that proper training is received and that all programs are run in the most effective way. Program implementation is a cycle that includes the adoption, delivery, uptake, and dissemination of a program. Because truly effective programs must be both designed well, and implemented skillfully, fidelity assessment are critical to evaluate the implementation process. Data linking fidelity to programmatic outcomes can facilitate sites' decisions to adopt programs based on their resources, organizational goals, and population. Fully understanding the impact treatment fidelity has on outcomes is a necessary step towards closing the gap between the interventions science deems effective, and those that produce optimal outcomes when implemented into the wider community.

References

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

Albarricin, D., Johnson, B., Fishbein, M., and Muellerleile, P. (2001). Theories of reasoned action and planned behavior as models of condom use: A meta-analysis. *Psychological Bulletin, 127,* pp. 142 – 161.

Avery-leaf, S. and Cascardi, M. (2002). Dating violence and education: Prevention and early Intervention strategies. In Schewe, P (Ed.): *Preventing violence in relationships: Interventions across the life span.* Washington, D.C.: American Psychological Association.

Ball, B., Kerig, P. K, and Rosenbluth, B. (in Press). "Like a family but better because you can actually trust each other": The Expect Respect Dating Violence Prevention Program for at-risk youth. Submitted to *Journal of Health Promotion and Practice.*

Battistich, V., Schaps, E., Watson, M., Solomon, D., & Lewis, C. (2000). Effects of the Child Development Project on students' drug use and other problem behaviors. *Journal of Primary Prevention, 21,* pp. 75 – 99.

Blakely, C.H., Emshoff, J.G., and Roitman, D.B. (1984). Implementing innovative programs In public sector organizations. *Applied Social Psychology, 5,* pp. 87 – 108.

Blakely, C.H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W., Roitman, D. B. and Emshoff, J. G. (1987). The fidelity-adaptation debate: implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15,* pp. 253 – 268.

Bond. G.R., Becker, D.R. and Drake, R.E. (1997). A fidelity scale for the individual placement And support model of supported employment. *Rehabilitation Counseling Bulletin, 40,* pp. 265 – 284.

Borelli, J.L. and David, D.H. (2004). Attachment theory as a guide to psychotherapy practice. *Imagination, Cognition, and Personality, 23,* pp. 257 – 287.

Brekke, J. S. and Wolkon, G. H. (December 1988). Monitoring program implementation in community mental health settings. *Evaluation and Health Professions, 11 (4),* pp. 425 - 440.

Bryk, A. S., Raudenbush, S. W., and Condon, R. (1996). *HLM: Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs.* Chicago, Illinois: Scientific Software.

Bryk, A. S. and Raudenbush, S. W. (2002). Hierarchical linear models, 2nd Edition. Sage: Thousand Oaks, California.

Calsyn, R.J. (2000). A checklist for critiquing fidelity studies. *Mental Health Services Research, 2,* pp. 107 – 113.

Centers for Disease Control and Prevention (2007). The effectiveness of universal school-based programs for the prevention of violent and aggressive behavior. *Morbidity and Mortality Weekly Report, 56,* pp. 1 – 12.

Cronbach, L.J., Yalow, E., and Schaeffer, G. (1980). A mathematical structure for analyzing Fairness in selection. *Personnel Psychology, 33,* pp. 693 – 704.

Dane, A.V. and Schneider, B.H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* pp. 23 – 24.

Dusenbury, L., Brannigan, R., Hansen, W. B., Walsh, J., and Falco, M. (November 2004). Quality of implementation: Developing measures crucial to understanding the diffusion of preventative interventions. *Health Education Research, 20 (3),* pp. 308 – 313.

Dusenbury, L., Brannigan, R., Falco, M., and Hansen, W.B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18 (2),* pp. 237 – 256.

Emshoff, J., Blakely, C., and Gray, D. (2003). An ESID case study at the federal level. *American Journal of Community Psychology, 32,* pp. 345 – 357.

Eysenck, H. (1984). Meta-analysis: An abuse of research integration. *Journal of Special Education, 18,* pp. 41 – 59.

Fishbein, M. and Yzer, M. C. (2003). Using theory to design effective health behavior interventions. *Communication Theory, 13,* pp. 164 – 183.

Foshee, V.A., Bauman, K.E., Greene, W., Koch, G., Linder, G.F., and MacDougall, J.E. (2000). The safe dates program: 1-year follow-up. *American Journal of Public Health, 90,* pp. 1619 – 1622.

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

Glasgow, R. E. and Emmons, K. M. (2007). How can we increase translation of research into practice? *Annual Review of Public Health, 28,* pp. 413 - 433.

Graham, I.D., Logan, J., Harrison, M.B., Straus, S.E., Tetroe, J., Caswell, W., and Robinson, N. (2006). Lost in knowledge translation: Time for a map? *The Journal of Continuing Education in the Health Professions, 26,* pp. 13 – 24.

Gresham, F.M., Gansle, K.A., Noell, G.H., Cohen, S., and Rosenblum, S. (1993). Treatment integrity of school-based behavioral intervention studies. *School Psychology Review, 22,* pp. 245 – 272.

Hall, G.E. and Loucks, S.F. (1978). *Innovation configurations: Analyzing the adaptation of innovations.* Paper presented at the meeting of the American Educational Research Association, Toronto.

Heck, R.H. and Thomas, S.L. (2000). *An introduction to multilevel modeling techniques.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hohmann, A.A. and Shear, M.K. (2002). Community based intervention research: Coping with the "noise" of real life in study design. *American Journal of Psychiatry, 159,* pp. 201 – 207.

Holtz, K. D. and Thwombly, E. C. (2007). A preliminary evaluation of the effects of a science education curriculum on changes on knowledge of drugs in youth. *Journal of Drug Education, 37,* pp. 313 – 333.

Hoyle, R.H., Harris, M.J., and Judd, C.M. (2001). *Research methods in social relations, 7th edition.* Belmont, California: Wadsworth Publishing.

Institute of Medicine. (2001). *Improving the quality of long-term care.* Washington D.C: National Academy Press.

Jaffe, P.G., Sunderman, M., and Reitzel, D. (1992). An evaluation of a secondary school primary prevention program on violence in intimate relationships. *Violence and Victims, 7,* pp. 129 – 146.

Kelly, J.A., Heckman, T.G., Stevenson. L.Y., and Williams, P.N. (2000). Transfer of research-based HIV prevention interventions to community service providers: Fidelity and adaptation. *AIDS Education and Prevention, 12,* pp. 87 – 98.

Klein, K. J. and Speer-Sorra, J. (1996). The challenge of innovation implementation. *Academy of Management Review, 21 (4),* pp. 1055 – 1080.

Mcgrew, J. H., Bond, G. R., Dietzen, L., and Salyers, M. (August 1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology, 62 (4),* pp. 670 – 678.

Miller, R. L. and Shinn, M. (June 2005). Learning from communities: Overcoming difficulties in dissemination of prevention and promotion efforts. *American Journal of Community Psychology, 35,* pp. 169 – 183.

Mowbray, C. T., Holter, M., Teague, G. B., and Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24 (3),*

pp. 315 – 340.

Mowbray, C. T., Bybee, D., Holter, M., & Lewandowski, L. (March 2006). Validation of fidelity rating instrument for consumer-operated services. *American Journal of Evaluation, 27 (1),* pp. 9 – 27.

Nabi, R. L., Southwell, B., and Hornik, R. (2002). Predicting intentions versus predicting behaviors: Domestic violence prevention from a theory of reasoned action perspective. *Health Communication, 14,* pp. 429 – 449.

Panzano, PC. And Roth, D. (2006). The decision to adopt evidence-based and other innovative mental health practices: Risky business? *Psychiatric Services, 57,* pp. 1153 – 1161.

Peterson, L., Homer, A.L., and Wonderlich, S.A. (1982). The integrity of independent variables on behavior analysis. *Journal of Applied Behavior Analysis, 13,* pp. 477 – 492.

Resnick, B., Bellg, A.J., and Borelli, B. (2005). Examples of implementation and evaluation of treatment fidelity in BCC studies: Where we are and where we need to go. *Annals of Behavioral Medicine, 29,* pp. 46 – 54.

Ridgely, M.S. and Jerrell, J.M. (1996). Analysis of three interventions for substance abuse treatment of severely mentally ill people. *Community Mental Health Journal, 32,* pp. 561 – 572.

Rogers, E.M. (1995). *Diffusion of Innovations.* Free Press; New York City, New York.

Rogers-Weise, M.R. (1992). A critical review of parent training research. *Psychology in Schools, 29,* pp. 229 – 236.

Rohrbach, L. A., Dent, C. W., Skara, S., Sun, P., and Sussman, S. (In Press). Fidelity of implementation in project Towards No Drub Abuse (TND): A comparison of classroom teachers and program specialists. To be published in the *Journal of Prevention Science.*

Safeplace. (2006). Expect Respect Safeplace school-based services: Program evaluation for the 2004-2005 academic year. Austin, Texas.

Shinn, M. (2003). Understanding implementation of programs in multi-level systems. *Prevention and Treatment, 6,* Item 22.

Smith, B. and Sechrest, L. (1991). Treatment of aptitude X treatment interactions. *Journal of Consulting and Clinical Psychology, 59,* pp. 233 – 244.

Summerfelt, W. S. (2003). Program strength and fidelity in evaluation. *Applied Developmental Science, 7 (2),* pp. 55 – 61.

Szulanski, G., and S.G. Winter (2002). Getting it right the second time," *Harvard Business Review*, *80,* pp. 62 – 69.

United States Department of Agriculture Food and Nutrition services (2007). National School Lunch Program Participation. Retrieved from http://www.fns.usda.gov/pd/cnpmain.htm

Unrau, Y.A. (2001). Using client interviews to illuminate outcome sin program logic models: A case example. *Evaluation and Program Planning, 24,* pp. 353 – 361.

Wade, D.T. (2001). Research into the lack box of rehabilitation: The risk of a type III error. *Clinical Rehabilitation, 15,* pp. 1 – 4.

Wandersman, A. (2003). Community science: Bridging the gap between science and practice with community-centered models. *American Journal of Community Psychology, 31,* pp. 227 – 242.

Weisz, A.N. and Black, B. (2001). Evaluating a sexual assault and dating violence prevention program for urban youths. *Social Work Research, 25,* pp. 89 – 102.

Appendix A

Facilitator Fidelity Interview

Safeplace, CDC, and Georgia State University are collaborating to study the processes by which Expect Respect Clubs are being implemented throughout Austin and Ohio. As part of that study, we are asking all group facilitators to participate in an interview.

There are no right and wrong answers to the interview questions. The purpose is to understand how facilitators are using the Expect/Respect Club model, not to determine who is and who is not doing a good job. Your answers do not have any positive or negative implications for you or your organization. Please be honest as you answer questions during the interview.

We are required to read you a consent form before beginning the interview. May I read it to you now?

Thank you very much for your help.

STAFF

What week of the program are you currently in?

Sex -

H - same sex single facilitator
M - co-facilitators (mixed or same sex)
L - opposite sex facilitators only

Race:

Experience –

H –experience facilitating teen groups
M – experience facilitating groups or working with teens, but not both
L – no teen or group experience

Education –

H - MSW or MA degree in a counseling-related field with appropriate licensure
M – Bachelors degree (or masters without licensure)
L – less than bachelors degree

Affiliation

H – External to school (to aid confidentiality)
M – School staff or faculty, with careful attention to student trust and rapport issues

L – Assigned school staff as a matter of convenience

Training

H-Facilitators received training from Expect Respect developers

M-Facilitators received training in using the Expect Respect materials from others

L-Facilitators have not received specific training for Expect Respect groups

TRAINING

- What aspects of the training you received for this program did you find helpful in running your program?
- What aspects of the training did you find unnecessary?
- Was it easy to translate things learned in training into program implementation?
- How important did you perceive the need to run your program like the original?
- Did you feel confident leaving the training sessions in your ability to implement the program?
- What else would you have wanted to get out of the training sessions that was not offered?

MATERIALS

H – Materials provided or suggested in the curriculum are used in all or almost all sessions

M – Materials provided or suggested in the curriculum are used some of the time

L – Provided materials rarely or never used

THEMES

H - Content or discussion about abusive or healthy relationships occurred in each session (relationships could be in families, dating relationships, friendships, work, etc.).

M - Content or discussion about abusive or healthy relationships occurred in most sessions.

L – Content or discussion about abuse or healthy relationships occurred in less than half of the sessions?

Please identify other content or themes discussed: _____.

How often were the following **topics** covered during your group sessions?

	Almost Every session	Most or Some sessions	One session	Never
Healthy relationships	**A.**	**B.**	**C.**	**D.**
Defining abuse and respect	**A.**	**B.**	**C.**	**D.**
Dating expectations	**A.**	**B.**	**C.**	**D.**
Consent	**A.**	**B.**	**C.**	**D.**
Healthy boundaries	**A.**	**B.**	**C.**	**D.**
Friendships	**A.**	**B.**	**C.**	**D.**
Power and control	**A.**	**B.**	**C.**	**D.**

Gender roles and stereotypes	**A.**	**B.**	**C.**	**D.**
Personal strength	**A.**	**B.**	**C.**	**D.**
Warning signs of abuse	**A.**	**B.**	**C.**	**D.**
Breaking up safely	**A.**	**B.**	**C.**	**D.**
Coping with stress	**A.**	**B.**	**C.**	**D.**
Helping others	**A.**	**B.**	**C.**	**D.**
Jealousy	**A.**	**B.**	**C.**	**D.**
Recognizing forms of violence in relationships	**A.**	**B.**	**C.**	**D.**
Reporting sexual harassment or dating violence to school authorities	**A.**	**B.**	**C.**	**D.**
Reporting dating violence to law enforcement	**A.**	**B.**	**C.**	**D.**

Which of these messages did you discuss in group?

	Agree	Not sure	Disagree
You are not alone and some of your peers have been through similar experiences.	**A.**	**B.**	**C.**
There are people available to help and support you.	**A.**	**B.**	**C.**
You deserve to be treated with respect in all your relationships.	**A.**	**B.**	**C.**
Violence and abuse are not part of healthy relationships.	**A.**	**B.**	**C.**
It's OK to express your feelings.	**A.**	**B.**	**C.**
Controlling behaviors are a warning sign of abuse.	**A.**	**B.**	**C.**
In healthy relationships partners share in making decisions and have equal powers.	**A.**	**B.**	**C.**
You can recover from abuse/violence in your life.	**A.**	**B.**	**C.**
Talking to other people about the abuse in your life can help you.	**A.**	**B.**	**C.**
Control and jealousy are not signs of love.	**A.**	**B.**	**C.**
Good communication is the key to healthy relationships.	**A.**	**B.**	**C.**
When someone respects you they accept you for who you are.	**A.**	**B.**	**C.**

IMPLEMENTATION

- How is the program working out for you and your staff?
- Are you finding it easy/difficult to follow through on each aspect of the program?
    - Which aspects?
    - Why is that?
- What aspects of the program are working well?
    - Why do you think that is?
- What aspects of this program are you finding it difficult to implement?
- Are there aspects of the program's implementation that you think are unnecessary? Why?
- Are there aspects of the program that you have intentionally avoided or left out?
    - Why?
- Have you implemented any changes/improvements to the existing program?
    - Can you describe them?
    - Were these modifications suggested or recommended by any program trainers?
        ...1. If not, did you consult with program trainers regarding the changes/modifications that your school was making?
    - How were decisions to modify the program made?
        ...1. Did modifications occur that were intended by program staff?
        ...2. Did modifications occur due to lack of resources (time, staff, money, etc.)?
        ...3. When were modifications made (before program implementation, after a couple sessions, etc.)?
- What are the unique needs or qualities of the youth (school climate, staff, local community) you work with that are beyond the reach of this program?
- What evidence do you have for the effectiveness of the program? (Teacher reports, student reports etc.)
    - Are the students active participants in the program?
        ...1. Do they seem to enjoy it?
    - Is the program supported by the faculty and staff?
    - Tell me more.

PROCESS

School liaison/contact person

H – a member of school staff or faculty – someone with knowledge and access to school logistics control

A – someone not connected with school (facilitator) – but with good relationships with school

U – someone without good relationship to school

Posters and/or information cards

H- in places where all students will see/receive them

M - in places where some students will see/receive them

L – no posters

Staff orientation

H – covers all or most of the information on the orientation outline

M – covers some of the information
L – covers little of the information or no orientation

Staff orientation
H – all or almost all faculty and staff attend
M – most faculty attend, but not staff
L – no orientation or very low attendance

Staff orientation

H – Includes Choose Respect video
L - does not include video

School agreement form
H – presented and signed
L – not used

Individual intake sessions –
H – held with all participants and covers all required topics (see manual)
M – not held with some participants or some topics not covered
L – not part of the process

District or school policy regarding sexual harassment
H – is well-defined, understood and followed
M – can be found, but may not be well-defined, understood, or followed
L – none in place

Intake form
H – Includes demographic information and reason for referral (or self-referral)
M – Is missing demographic information or reason for referral
L – no intake form

Program sessions were held
H – During school hours, either rotating hours to avoid students missing core subjects repeatedly or during a planning/study hall-like period
M – During school hours, requiring students to miss core subjects repeatedly
L – After school or during lunch

Session duration
H – 50 minutes or more
M – 30-50 minutes
L – less than 30 minutes

Session frequency
H – weekly except for holidays and testing

M – two or three times per month
L – no regular schedule

Group size
H – 6-10
M – Under or over by 1 or 2
L – under or over by more than 2

Activities *
H – Activities provided or suggested in the curriculum used all or almost all of the time
M – Activities provided or suggested used some of the time
L – Activities provided or suggested rarely or never used

Which of these **activities** happened in your group?

Activity	We definitely did this in group	We did not do this in group
We viewed a video on dating violence.	**A.**	**B.**
Our group met with a group of the opposite sex.	**A.**	**B.**
We did a project to educate other people about dating violence.	**A.**	**B.**

Discussion questions suggested in the curriculum
H – or similar questions used in all or most sessions
M –or similar questions used in some sessions
L – or similar questions rarely or never used

Use of group time
H-The facilitator frequently uses group time (beyond check in) to acknowledge, listen to and support individual members who are in distress or sharing feelings.
M-The facilitator occasionally uses group time (beyond check in) to acknowledge, listen to and support individual members.
L-The facilitator uses group time strictly for the curriculum or other structured activities.

Use of group time (flexibility)
H- The facilitator frequently gives room for group members to raise questions and issues (or initiate discussion) about relationships that are relevant to them even if they are not planned in the curriculum.
M – The facilitator occasionally gives room to group members to raise questions and issues about relationships even if they are not planned in the curriculum.

L – The facilitator follows the weekly curriculum

Facilitator interaction with school contact person
H – Meets regularly with school contact person before or after group sessions
M – meets occasionally with school contact person
L – meets with contact person rarely or never

If co-facilitators*
Met and planned prior to sessions
H – all or almost all of the time
M – some of the time
N – rarely or never

If co-facilitators *
Met, de-briefed and provided feedback after sessions
H – all or almost all of the time
M – some of the time
L – rarely or never

Individual follow-up sessions, crisis intervention or referrals made
H – always used when needed
M – sometimes used when needed
L - rarely or never used (unless never needed)

PRIORITIZATION OF REFERRALS
H – All referrals assessed and these assessments used to determine appropriateness for inclusion in the group
M – Some students included without assessment

Privacy Conditions
1)no sign on door, no other means of identifying group
2)no unblocked window into group room
3)location conducive to privacy (for example, not adjacent to Principal's office where voices would be heard in next room")

H – All three conditions met during session
M – One or two conditions met during session
L - None of the conditions met

Student pass
H – systematic easy way of getting students excused from class, such as the sample hall pass
M – inconsistent or somewhat confusing system for class excuse
L – no system in place

Confidentiality policy, including exceptions
H – Explained to all participants
M – Explained to some or most participants OR exceptions not explained
L – No policy in place

Parental consent
H – issues of consent identified and system in place
M – issues identified, system not in place or not well implemented
L – issues not identified or major problems in implementation of a consent process

Parental consent required by school and/or state law
Parental consent not required

Obtaining parental consent
H - Procedures in place and well implemented
M – Procedures in place but obtaining consent is difficult (or an obstacle)
L – No clear procedures in place/or major problems in obtaining consent

Referral process
H – system in place for faculty/staff referrals and self-referral
M – either faculty/staff or self referral process not in place
L – no clear process for referring students to the program

Program Participants
H – all participants have experienced, witnessed or perpetrated domestic, sexual, or dating violence
M – some participants are at risk for violence or have an interest in these issues
L – some participants are just interested in getting out of class or being with friends

Terms of participation
H – voluntary
M – some coercion
L – mandatory for certain participants

GROUP MEMBERSHIP
First Half Of Sessions
H - The group starts out full or mostly full and a few new members are added in subsequent weeks.
M - The group starts out small and new members join in subsequent weeks until the group is full
L- A core group is never established, new members join and then leave

Second Half Of Sessions

H – no new members after the halfway point
M – one or two exceptions to this rule
L – no boundaries to entrance and exit from the group – totally open

GROUP PROCESS

Check-in or icebreaker at beginning of sessions
H – used for every session
M – used inconsistently
L – not used for sessions

Session Content & Agenda
H – Session content always covered as planned from curriculum, including sufficient didactic, discussion, exercises as indicated
M – Session content usually covered, but some inconsistency
L – Very little correspondence between curriculum and session content

General Support
H – 10-15 minutes spent on general support, review of personal issues and experiences during sessions
M – Less than 10 minutes given to general support in multiple sessions
L – Less than 5 minutes given to general support or more than 15 minutes consistently given to general support

Participation
H – Participation is balanced (everyone participates more or less equally)
M – Participation is semi-balanced, but some participants may dominate while others are often withdrawn
L – Participation is very much dominated by a few participants

Participant Support
H – a high level of support among participants
M – a moderate level of support among participants
L – a low level of support among participants

Facilitator Role
H - Over time the facilitator talks less and allows greater management of group time by group members
M - The facilitator plays a steady role in facilitating sessions over time
L - The facilitator leads the group by doing most of the talking

EVALUATION

Pre-post survey
H – all or most of the participants complete pre and post surveys of knowledge attitudes and behavior

M – Some, but not a majority of participants completed either pre and post surveys
L – no survey process in place

Group attendance and process log
H – used consistently, completely and accurately
M – some lack of consistency, completion, or accuracy
L – not used

Student feedback form
H – completed by everyone (or almost everyone)
M – completed by some students
L – completed by very few or no students

School contact person feedback form
H – completed
L – not completed

Focus groups
H – implemented with good participation and representation (e.g., both boys and girls)
M – implemented, but less than ideal participation or representation
L – not implemented

Is there an underlying theme or philosophy that you feel serves as the foundation for the entire Expect Respect program? If so, how would you state it?

TECHNICAL ASSISTANCE

- Did you receive any technical assistance?
  - What type of assistance did you receive (Phone call personal visit etc.)?
  - Was technical assistance offered or did you seek it?
  - How often was technical assistance offered /utilized?
  - How available were program trainers to your site for technical assistance?
  - How helpful was the technical assistance?
    - What could have been more helpful?

Appendix B

Pre-post Student Survey

Instructions

We would like to ask you a few questions about your relationships. This is not a test, so there are no right or wrong answers.

Please read the directions and questions carefully and answer the questions as honestly as you can. The Expect Respect Counselor will assist you if you need help.

Your answers are completely confidential. School personnel will not see your questionnaire. By answering the questions you help us to understand what is important to teens like you and improve the Expect Respect Groups.

Part 1: About You

1. Name _____

2. Date _____ **3. School** _____

4. What grade are you in?

❑ 1 6th ❑ 2 7th ❑ 3 8th ❑ 4 9th ❑ 5 10th ❑ 6 11th ❑ 7 12th

5. Are you

❑ 1 Female ❑ 2 Male

6. Which of the following best describes you? (Mark one or more.)

❑ 1 American Indian or Alaska Native
❑ 2 Asian
❑ 3 Black or African American
❑ 4 Hispanic or Latino
❑ 5 Native Hawaiian or Other Pacific Islander
❑ 6 White
❑ 7 Other

7. Is this your first year in group?

❑ 1 Yes ❑ 2 No If not, how many years **before this year** have you been in group? ❑ 1 1 ❑ 2 2 ❑ 3 3

8. How did you hear about the group? _____

Part 2: Are the following statements true for you ?

	Very True	True	Not sure	False	Very False
1. There are people I can depend on to help me if I really need to.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
2. If something went wrong, no one would come to my assistance.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
3. There is an adult I would want to talk to about important decisions in my life.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
4. If I were having problems I could turn to a trustworthy adult.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
5. I cannot depend on my peers for help.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
6. I feel comfortable talking about my problems with at least one of my friends.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
7. I have friends I can count on in an emergency.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
8. There is someone in my life who cares about my feelings.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5

Part 3: Now we will ask you some questions about your dating relationships. By dating relationship we're talking about having a boyfriend or a girlfriend, or hanging out in a romantic way, or hooking up with one person.

9. Have you ever had a boyfriend or girlfriend?

❑ 1 Yes ❑ 2 No

10. Have you had a boyfriend or girlfriend in the last three months?
❑ 1 Yes ❑ 2 No

No matter how well people get along there are times when they disagree, get annoyed, or have fights. When you were upset or angry at your boyfriend or girlfriend in the past three months,
how often did any of these things happen?
If you did not have a dating partner, please think about a close friend.

	Often	**Sometimes**	**Rarely**	**Never**
11. I **tried to make this person feel jealous.**	❑ **1**	❑ **2**	❑ **3**	❑ **4**
12. This person **tried to make me feel jealous.**	❑ **1**	❑ **2**	❑ **3**	❑ **4**
13. I **told this person how I felt and why.**	❑ **1**	❑ **2**	❑ **3**	❑ **4**
14. This person **told me how he/she felt and why.**	❑ **1**	❑ **2**	❑ **3**	❑ **4**
15. I **made fun of this person in front of others.**	❑ **1**	❑ **2**	❑ **3**	❑ **4**
16. This person **made fun of me in front of others.**	❑ **1**	❑ **2**	❑ **3**	❑ **4**
17. I **threatened to hurt this person.**	❑ **1**	❑ **2**	❑ **3**	❑ **4**

18. This person **threatened to hurt me.**	❑ 1	❑ 2	❑ 3	❑ 4
19. I **called this person mean names.**	❑ 1	❑ 2	❑ 3	❑ 4
20. This person **called me mean names.**	❑ 1	❑ 2	❑ 3	❑ 4
21. I **hit, kicked or punched something, like a wall or table.**	❑ 1	❑ 2	❑ 3	❑ 4
22. This person **hit, kicked or punched something, like a wall or table.**	❑ 1	❑ 2	❑ 3	❑ 4
23. I **left the room to cool down.**	❑ 1	❑ 2	❑ 3	❑ 4
24. This person **left the room to cool down.**	❑ 1	❑ 2	❑ 3	❑ 4
25. I **pushed, shoved or shook the person.**	❑ 1	❑ 2	❑ 3	❑ 4
26. This person **pushed, shoved or shook me.**	❑ 1	❑ 2	❑ 3	❑ 4
27. I **yelled or screamed at this person.**	❑ 1	❑ 2	❑ 3	❑ 4
28. This person **yelled or screamed at me.**	❑ 1	❑ 2	❑ 3	❑ 4
29. I **calmed myself down before I talked to this person.**	❑ 1	❑ 2	❑ 3	❑ 4

30. This person **calmed themselves down before they talked to me.**	❑ 1	❑ 2	❑ 3	❑ 4
31. I **gave this person the silent treatment.**	❑ 1	❑ 2	❑ 3	❑ 4
32. This person **gave me the silent treatment.**	❑ 1	❑ 2	❑ 3	❑ 4
33. I **asked this person what he/she was feeling.**	❑ 1	❑ 2	❑ 3	❑ 4
34. This person **asked me what I was feeling.**	❑ 1	❑ 2	❑ 3	❑ 4
35. I **let this person know what was important to me.**	❑ 1	❑ 2	❑ 3	❑ 4
36. This person **let me know what was important to him/her.**	❑ 1	❑ 2	❑ 3	❑ 4

Part 4: How do you feel about yourself?

Are the following statements true for you?	**Very true**	**True**	**Not Sure**	**False**	**Very false**
37. I really don't like myself very much.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
38. I sometimes feel so bad about myself that I wish I were somebody else.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
39. I usually feel I'm the kind of person I want to be.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
40. I feel I can do things as well as other people can.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5

41. I feel that I am a special or important person.	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5

How confident are you that you can do the following things?

	Very confident	Confident	Somewhat confident	Not confident
42. Stand up for yourself in your relationships.	❑ 1	❑ 2	❑ 3	❑ 4
43. Break up with somebody who is putting you down.	❑ 1	❑ 2	❑ 3	❑ 4
44. Control your temper when you are angry.	❑ 1	❑ 2	❑ 3	❑ 4
45. Say "no" to doing something sexual when you are not ready, even though your partner keeps begging.	❑ 1	❑ 2	❑ 3	❑ 4
46. Tell your dating partner when you feel hurt or upset with him or her.	❑ 1	❑ 2	❑ 3	❑ 4
47. Choose a dating partner who will treat you well.	❑ 1	❑ 2	❑ 3	❑ 4
48. Talk with your dating partner about sex to find out whether they are ready.	❑ 1	❑ 2	❑ 3	❑ 4

Part 5: What do you do when you feel worried or upset?

In the past three months, when you faced difficulties, felt tense or remembered painful situations how often did you do any of the following?	Often	Sometimes	Rarely	Never
49. Skipped school.	❑ 1	❑ 2	❑ 3	❑ 4
50. Listened to music.	❑ 1	❑ 2	❑ 3	❑ 4
51. Wrote poetry, songs or in your journal.	❑ 1	❑ 2	❑ 3	❑ 4
52. Tried to get away from people by staying in your room.	❑ 1	❑ 2	❑ 3	❑ 4
53. Went to the gym, shot hoops, played football, lifted weights, etc.	❑ 1	❑ 2	❑ 3	❑ 4
54. Ate a lot even when you felt full.	❑ 1	❑ 2	❑ 3	❑ 4
56. Cut, burned or hurt yourself in other ways.	❑ 1	❑ 2	❑ 3	❑ 4
57. Talked to family members.	❑ 1	❑ 2	❑ 3	❑ 4
58. Tried to be funny and not take it seriously.	❑ 1	❑ 2	❑ 3	❑ 4
59. Talked to a teacher or counselor	❑ 1	❑ 2	❑ 3	❑ 4
60. Got angry and yelled at people.	❑ 1	❑ 2	❑ 3	❑ 4
61. Tried to be close with people who care about you.	❑ 1	❑ 2	❑ 3	❑ 4
62. Drank alcohol or used drugs.	❑ 1	❑ 2	❑ 3	❑ 4
63. Talked to a friend.	❑ 1	❑ 2	❑ 3	❑ 4

Part 6: Which of the following behaviors would you consider to be abusive?

If someone ……	Definitely not abuse	Probably not abuse	Not sure	Probably abuse	Definitely abuse
64. Acts like they own you	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
65. Wants to be in charge of everything that goes on	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
66. Wants you to spend all your time with them and drop other friends	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
67. Accepts your opinions	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
68. Puts you down and calls you mean names	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
69. Drives fast to scare you	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
70. Pressures you to dress a certain way	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
71. Accepts when you choose to spend time alone or with your family	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
72. Controls you	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
73. Says they will commit suicide if you break up	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
74. Slaps, pushes or shoves you	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5
75. Follows you around when you told him/her not to	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5

Part 7: Being in group has helped me…

	Agree	*Not Sure*	*Disagree*
76. Communicate my thoughts and feelings	1	2	3
77. Increase my personal safety	1	2	3
78. Learn different ways of coping with my problems	1	2	3
79. Feel supported	1	2	3
80. Stand up for myself	1	2	3
81. Increase my understanding of abusive and respectful relationships	1	2	3
82. Increase my skills for healthy relationships	1	2	3
83. Know how to help a friend in an abusive relationship	1	2	3
84. Speak up when I see abusive or harassing behavior	1	2	3

Part 8: Questions About Group Leaders.
If your group has more than one leader, please think about the one you feel closest to.

	Not at all	**Somewhat**	**A lot**
85. I trust my group leader.	1	2	3

86. I feel respected by my group leader.	**1**	**2**	**3**
87. I respect my group leader.	**1**	**2**	**3**
88. I think I am a lot like my group leader.	**1**	**2**	**3**
89. I admire my group leader.	**1**	**2**	**3**
90. My group leader understands the things I'm going through.	**1**	**2**	**3**
91. I see my group leader as a friend.	**1**	**2**	**3**
92. I see my group leader as a teacher.	**1**	**2**	**3**
93. I see my group leader as a role model.	**1**	**2**	**3**
94. How much did the group help you feel better about yourself?	**1**	**2**	**3**
	Not at all	Somewhat	A lot

95. How much did you feel accepted by the group?	**1**	**2**	**3**
96. How much did you feel you could trust the group?	**1**	**2**	**3**
97. My group leader likes me.	**1**	**2**	**3**
98. My group leader shows me that he/she cares about me.	**1**	**2**	**3**

99. What is the most important thing you learned in this group?

100. Has your life changed at all by being in this group? If yes, how?

101. What would you change about this group?

Thanks for your help in improving this program!