Effects of bystander programs on the prevention of sexual assault among adolescents and college students: a systematic review

Heather Hensman Kettrey, Robert A. Marx, and Emily E. Tanner-Smith
The Campbell Library comprises:

- Systematic reviews (titles, protocols and reviews)
- Policies and Guidelines Series
- Methods Series

Go to the library to download these resources, at:
www.campbellcollaboration.org/library/

Better evidence for a better world
<table>
<thead>
<tr>
<th>Colophon</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
</tbody>
</table>
| **Authors** | Heather Hensman Kettrey, PhD  
Robert A. Marx, MA, MS  
Emily E. Tanner-Smith, PhD |
| **DOI** | 10.4073/csr.2019.1 |
| **No. of pages** | 156 |
| **Citation** | Kettrey, H. H., Marx, R. A. & Tanner-Smith, E. E. Effects of bystander programs on the prevention of sexual assault among adolescents and college students: a systematic review. Campbell Systematic Reviews 2019:1  
DOI: https://10.4073/csr.2019.1 |
| **ISSN** | 1891-1803 |
| **Copyright** | © Kettrey et al.  
This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. |
| **Roles and responsibilities** | Dr Kettrey, the lead review author, coordinated the review team and assumed responsibility for the implementation of the project throughout its duration. Specific tasks included compiling the sample of research reports, creating the database, coding studies, analyzing data, and preparing the Campbell review. Mr Marx, the second review author, collaborated closely with Dr Kettrey to compile the sample of research reports, code studies, and make methodological decisions throughout the duration of the project. Dr Tanner-Smith, the third review author, provided methodological guidance and mentorship to Dr Kettrey and Mr Marx throughout all phases of data collection and analysis. Dr Kettrey anticipates updating the review at least once every five years, pending continual accrual of sufficient research on the topic. |
| **Editors for this review** | Editor: Charlotte Gill  
Managing editor: Liz Eggins |
| **Sources of support** | This review was supported by a grant from the Campbell Collaboration (CSR1.60). |
| **Declarations of interest** | The authors have no conflicts of interests to report. |
| **Corresponding author** | Heather Hensman Kettrey, PhD  
Department of Sociology, Anthroplogy, and Criminal Justice  
132 Brackett Hall  
Clemson University  
Clemson, South Carolina 29634-1356  
E-mail: hkettre@clemson.edu  
Full list of author information is available at the end of the article. |
The Campbell Collaboration was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. Campbell offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of Campbell’s editors, librarians, methodologists and external peer reviewers contribute.

The Campbell Collaboration
P.O. Box 222 Skøyen
0213 Oslo, Norway
www.campbellcollaboration.org
# Table of contents

**PLAIN LANGUAGE SUMMARY**  
5  
**EXECUTIVE SUMMARY/ABSTRACT**  
7  
  Background  
  7  
  Objectives  
  8  
  Search methods  
  8  
  Selection criteria  
  8  
  Data collection and analysis  
  10  
  Results  
  11  
  Authors’ conclusions  
  12  
**BACKGROUND**  
14  
  The problem, condition or issue  
  14  
  The intervention  
  15  
  How the intervention might work  
  16  
  Why it is important to do the review  
  18  
**OBJECTIVES**  
20  
  The problem, condition or issue  
  20  
**METHODS**  
21  
  Criteria for considering studies for this review  
  21  
  Search methods for identification of studies  
  24  
  Data collection and analysis  
  26  
**RESULTS**  
30  
  Description of studies  
  30  
  Risk of bias in included studies  
  33  
  Synthesis of results  
  34  
**DISCUSSION**  
72  
  Summary of main results  
  72  
  Overall completeness and applicability of evidence  
  75  
  Quality of the evidence  
  76  
  Limitations and potential biases in the review process  
  76
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreements and disagreements with other studies or reviews</td>
<td>76</td>
</tr>
<tr>
<td><strong>AUTHORS’ CONCLUSIONS</strong></td>
<td>78</td>
</tr>
<tr>
<td>Implications for practice and policy</td>
<td>78</td>
</tr>
<tr>
<td>Implications for research</td>
<td>78</td>
</tr>
<tr>
<td><strong>REFERENCES</strong></td>
<td>80</td>
</tr>
<tr>
<td>References to included studies</td>
<td>80</td>
</tr>
<tr>
<td>References to excluded studies</td>
<td>83</td>
</tr>
<tr>
<td>References to studies awaiting classification</td>
<td>90</td>
</tr>
<tr>
<td>References to ongoing studies</td>
<td>90</td>
</tr>
<tr>
<td>Additional references</td>
<td>91</td>
</tr>
<tr>
<td><strong>INFORMATION ABOUT THIS REVIEW</strong></td>
<td>97</td>
</tr>
<tr>
<td>Review authors</td>
<td>97</td>
</tr>
<tr>
<td>Roles and responsibilities</td>
<td>98</td>
</tr>
<tr>
<td>Sources of support</td>
<td>98</td>
</tr>
<tr>
<td>Declarations of interest</td>
<td>98</td>
</tr>
<tr>
<td>Plans for updating the review</td>
<td>99</td>
</tr>
<tr>
<td><strong>FIGURES AND TABLES</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>APPENDIX: CODEBOOK</strong></td>
<td>128</td>
</tr>
<tr>
<td><strong>ELIGIBILITY CRITERIA</strong></td>
<td>129</td>
</tr>
<tr>
<td><strong>FULL-TEXT CODING</strong></td>
<td>132</td>
</tr>
<tr>
<td>Study level</td>
<td>132</td>
</tr>
<tr>
<td>Intervention and comparison groups</td>
<td>142</td>
</tr>
<tr>
<td>Outcomes</td>
<td>149</td>
</tr>
<tr>
<td>Effect sizes</td>
<td>151</td>
</tr>
</tbody>
</table>
Plain language summary

Bystander programmes increase bystander intervention but no effect on perpetrating sexual assault

Bystander sexual assault prevention programs have beneficial effects on bystander intervention but there is no evidence of effects on sexual assault perpetration. Effects on knowledge and attitudes are inconsistent across outcomes.

What is this review about?

Sexual assault is a significant problem among adolescents and college students across the world. One promising strategy for preventing these assaults is the implementation of bystander sexual assault prevention programs, which encourage young people to intervene when witnessing incidents or warning signs of sexual assault. This review examines the effects bystander programs have on knowledge and attitudes concerning sexual assault and bystander behavior, bystander intervention when witnessing sexual assault or its warning signs, and participants’ rates of perpetration of sexual assault.

What is the aim of this review?

This Campbell systematic review examines the effects of bystander programs on knowledge and attitudes concerning sexual assault and bystander intervention, bystander intervention when witnessing sexual assault or its warning signs, and the perpetration of sexual assault. The review summarizes evidence from 27 high-quality studies, including 21 randomized controlled trials.

What are the main findings of this review?

What studies are included?

This review includes studies that evaluate the effects of bystander programs for young people on (1) knowledge and attitudes concerning sexual assault and bystander intervention, (2) bystander intervention behavior when witnessing sexual assault or its warning signs, and (3) perpetration of sexual assault. Twenty-seven studies met the inclusion criteria. These included studies span the
period from 1997 to 2017 and were primarily conducted in the USA (one study was conducted in Canada and one in India). Twenty-one of the studies were randomized controlled trials and six were high quality quasi-experimental studies.

**Do bystander programs have an effect on knowledge/attitudes, on bystander intervention, or on sexual assault perpetration?**

Bystander programs have an effect on knowledge and attitudes for some outcomes. The most pronounced beneficial effects are on rape myth acceptance and bystander efficacy outcomes. There are also delayed effects (i.e., one to four months after the intervention) on taking responsibility for intervening/acting, knowing strategies for intervening, and intentions to intervene outcomes. There is little or no evidence of effects on gender attitudes, victim empathy, date rape attitudes, and on noticing sexual assault outcomes.

Bystander programs have a beneficial effect on bystander intervention. There is no evidence that bystander programs have an effect on participants’ rates of sexual assault perpetration.

**What do the findings of this review mean?**

The United States 2013 Campus Sexual Violence Elimination (SaVE) Act requires post-secondary educational institutions participating in Title IX financial aid programs to provide incoming college students with sexual violence prevention programming that includes a component on bystander intervention.

Bystander programs have a significant effect on bystander intervention. But there is no evidence that these programs have an effect on rates of sexual assault perpetration. This suggests that bystander programs may be appropriate for targeting the behavior of potential bystanders but may not be appropriate for targeting the behavior of potential perpetrators.

Beneficial effects of bystander programs on bystander intervention were diminished by six months post-intervention. Thus, booster sessions may be needed to yield any sustained effects.

There are still important questions worth further exploration. Namely, more research is needed to investigate the underlying causal mechanisms of program effects on bystander behavior (e.g., to model relationships between specific knowledge/attitude effects and bystander intervention effects), and to identify the most effective types of bystander programs (e.g., using randomized controlled trials to compare the effects of two alternate program models). Additionally, more research is needed in contexts outside of the USA so that researchers can better understand the role of bystander programs across the world.

**How up-to-date is this review?**

The review authors searched for studies up to June 2017. This Campbell systematic review was submitted in October 2017, revised in October 2018, and published January 2019.
Executive summary/Abstract

Background

Sexual assault among adolescents and college students

Sexual assault is a significant problem among adolescents and college students in the United States and globally. Findings from the Campus Sexual Assault study estimated that 15.9% of college women had experienced attempted or completed sexual assault (i.e., unwanted sexual contact that could include sexual touching, oral sex, intercourse, anal sex, or penetration with a finger or object) prior to entering college and 19% had experienced attempted or completed sexual assault since entering college (Krebs, Lindquist, Warner, Fisher, & Martin, 2009). Similar rates have been reported in Australia (Australian Human Rights Commission, 2017), Chile (Lehrer, Lehrer, & Koss, 2013), China (Su, Hao, Huang, Xiao, & Tao, 2011), Finland (Bjorklund, Hakkanen-Nyholm, Huttunen, & Kunttu, 2010), Poland (Tomaszew ska & Krahe, 2015), Rwanda (Van Decraen, Michielsen, Herbots, Rossem, & Temmerman, 2012), Spain (Vazquez, Torres, Otero, 2012) and in a global survey of countries in Africa, Asia, and the Americas (Pengpid & Peltzer, 2016).

The bystander approach

One promising strategy for preventing sexual assault among adolescents and young adults is the implementation of bystander programs, which encourage young people to intervene when witnessing incidents or warning signs of sexual assault. Bystander programs seek to sensitize young people to warning signs of sexual assault, create attitudinal changes that foster bystander responsibility for intervening (e.g., creating empathy for victims), and build requisite skills and knowledge of tactics for taking action (Banyard, 2011; Banyard, Plante, & Moynihan, 2004; Burn, 2009; McMahon & Banyard, 2012). Many of these programs are implemented with large groups of adolescents or college students in the format of a single training/education session (e.g., as part of college orientation). However, some programs use broader implementation strategies, such as advertising campaigns that post signs across college campuses to encourage students to act when witnessing signs of violence.

By treating young people as potential allies in preventing sexual assault, bystander programs have the capacity to be less threatening than traditional sexual assault prevention programs, which tend to address young people as either potential perpetrators or victims of sexual violence (Burn, 2009; Messner, 2015; [Jackson] Katz, 1995). Instead of placing emphasis on how young people may modify their individual behavior to either respect the sexual boundaries of others or reduce their
personal risk for being sexually assaulted, bystander programs aim to foster prerequisite knowledge and skills for intervening on behalf of potential victims. Thus, by treating young people as part of the solution to sexual assault, rather than part of the problem, bystander programs may limit the risk of defensiveness or backlash among participants (e.g., decreased empathy for victims, increased rape myth acceptance) (Banyard et al., 2004; Katz, 1995).

Objectives

The overall objective of this systematic review and meta-analysis was to examine what effects bystander programs have on preventing sexual assault among adolescents and college students. More specifically, this review addressed three objectives.

1. The first objective was to assess the overall effects (including adverse effects), and the variability of the effects, of bystander programs on adolescents’ and college students’ attitudes and behaviors regarding sexual assault.

2. The second objective was to explore the comparative effectiveness of bystander programs for different profiles of participants (e.g., mean age of the sample, education level of the sample, proportion of males/females in the sample, proportion of fraternity/sorority members in the sample, proportion of athletic team members in the sample).

3. The third objective was to explore the comparative effectiveness of different bystander programs in terms of gendered content and approach (e.g., conceptualizing sexual assault as a gendered or gender-neutral problem, mixed- or single-sex group implementation).

Search methods

Candidate studies were identified through searches of electronic databases, relevant academic journals, and gray literature sources. Gray literature searches included contacting leading authors and experts on bystander programs to identify any current/ongoing research that might be eligible for the review, screening bibliographies of eligible studies and relevant reviews to identify additional candidate studies, and conducting forward citation searches (searches for reports citing eligible studies) using the website Google Scholar.

Selection criteria

To be included in the review studies had to meet eligibility criteria in the following domains: Types of study, Types of participants, Types of interventions, Types of outcome measures, Duration of follow-up, and Types of settings.

Types of studies

To be eligible for inclusion in the review, studies must have used an experimental or controlled quasi-experimental research design to compare an intervention group (i.e., students assigned to a bystander program) with a comparison group (i.e., students not assigned to a bystander program).
Types of participants
The review focused on studies that examined outcomes of bystander programs that target sexual assault and were implemented with adolescents and/or college students in educational settings. Eligible participants included adolescents enrolled in grades 7 through 12 and college students enrolled in any type of undergraduate postsecondary educational institution. The mean age of samples could be no less than age 12 and no greater than age 25.

Types of interventions
Eligible intervention programs were those that approached participants as allies in preventing and/or alleviating sexual assault among adolescents and/or college students. Some part of the program had to focus on ways that cultivate willingness for a person to respond to others who are at risk for sexual assault. All delivery formats were eligible for inclusion (e.g., in-person training sessions, video programs, web-based training, advertising/poster campaigns). There were no intervention duration criteria for inclusion.

Eligible comparison groups must have received no intervention services targeting bystander attitudes/behavior or sexual assault.

Types of outcome measures
We included studies that measured the effects of bystander programs on at least one of the following primary outcome domains:

1. General attitudes toward sexual assault and victims (e.g., victim empathy, rape myth acceptance).
2. Prerequisite skills and knowledge for bystander intervention as defined by Burn (2009) (e.g., noticing sexual assault or its warning signs, identifying a situation as appropriate for intervention, taking responsibility for acting/intervening, knowing strategies for helping/intervening).
3. Self-efficacy with regard to bystander intervention (e.g., respondents’ confidence in their ability to intervene).
4. Intentions to intervene when witnessing instances or warning signs of sexual assault.
5. Actual intervention behavior when witnessing instances or warning signs of sexual assault.
6. Perpetration of sexual assault (i.e., participants' rates of perpetration).

Duration of follow-up
Studies reporting follow-ups of any duration were eligible for inclusion. When studies reported outcomes at more than one follow-up wave, each wave was coded and identified by its reported duration. Follow-ups of similar durations were analyzed together.
Types of settings
The review focused on studies that examined outcomes of bystander programs that target sexual assault and were implemented with adolescents and/or college students in educational settings. Eligible educational settings included secondary schools (i.e., grades 7-12) and colleges or universities. There were no geographic limitations on inclusion criteria. Research conducted in any country was eligible.

Data collection and analysis

Selection of studies
Once candidate studies were identified, two reviewers independently screened each study title and abstract for eligibility; disagreements between reviewers were resolved by discussion and consensus. Potentially eligible studies were then retrieved in full text, and these full texts were reviewed for eligibility, again using two independent reviewers.

Data extraction and management
Two reviewers independently double-coded all included studies, using a piloted codebook. Coding disagreements were resolved via discussion and consensus. The primary categories for coding were as follows: participant demographics and characteristics (e.g., age, gender, education level, race/ethnicity, athletic team membership, fraternity/sorority membership); intervention setting (e.g., state, country, secondary or post-secondary institution, mixed- or single-sex group); study characteristics (e.g., attrition, duration of follow-up, study design, participant dose, sample N); outcome construct (e.g., type, description of measure); and outcome results (e.g., timing at measurement, baseline and follow-up means and standard deviations or proportions).

Measures of treatment effect
During the coding process, relevant summary statistics (e.g., means and standard deviations, proportions, observed sample sizes) were extracted from research reports to calculate effect sizes. Effect sizes were reported as standardized mean differences (SMD), adjusted for small sample size (Hedges’ g). Positive effect size values (i.e., greater than 0) indicate a beneficial outcome for the bystander intervention group.

Data synthesis
Intervention effects for each outcome construct were synthesized via meta-analyses using random-effects inverse variance weights. All statistical analyses were conducted with the metafor package in R. Synthesis results are displayed using forest plots. Mean effect sizes are reported with their 95% confidence intervals.
Results

Objective 1: Effects on Knowledge, Attitudes, and Behavior

Knowledge/Attitudes. Effects for knowledge and attitude outcomes varied widely across constructs. The most pronounced beneficial effect in this domain was on rape myth acceptance. The effect for this outcome was immediate and sustained across all reported follow-up waves (i.e., from immediate post-test to six- to seven-months post-intervention). Intervention effects on bystander efficacy were also fairly pronounced, with an effect observed at both immediate post-test and one- to four-months post-intervention. A significant effect was not observed for this outcome six months post-intervention; however, this should be interpreted with caution, as only one study reported bystander efficacy effects at this follow-up period.

Effects on other knowledge and attitude outcomes were either delayed or unobserved. Intervention effects on taking responsibility for intervening or acting, knowing strategies for intervening, and intentions to intervene were non-significant at immediate post-test, but significant and beneficial by one- to four-months post-intervention. We found limited or no evidence of significant intervention effects on gender attitudes, victim empathy, date rape attitudes, and noticing sexual assault.

Behavior. The results indicated that bystander programs have a beneficial effect on bystander intervention behavior. However, this effect, which was observed at one-to four-months post-intervention, was not statistically significant at six months post-intervention. Bystander programs did not have a significant effect on sexual assault perpetration.

Objective 2: Effects for Different Participant Profiles

We had planned to conduct moderator analyses to assess a wide range of participant characteristics as potential effect size moderators. However, our review only yielded a sufficient number of studies \((n \geq 10)\) to conduct moderator analyses for the bystander intervention outcome domain. The results indicated that mean age, education level, and proportion of males/female were not statistically significant predictors of the magnitude of intervention effects.

Objective 3: Effects Based on Gendered Content/Implementation

We conducted moderator analyses to assess any differential effects of bystander programs on measured outcomes based on (1) the gender of perpetrators and victims in specific bystander programs and (2) whether programs were implemented in mixed- or single-sex settings. Our review only produced a sufficient number of studies \((n \geq 10)\) to conduct such moderator analyses for the bystander intervention outcome domain. We found that neither of these measures was a significant predictor of the effectiveness of bystander programs on bystander intervention.
Implications for practice and policy

The overwhelming majority of eligible studies assessing the effects of bystander programs were conducted in the United States. This is not necessarily surprising considering that the United States has implemented public policy that encourages the implementation of such programs on college campuses. The United States 2013 Campus Sexual Violence Elimination (SaVE) Act requires post-secondary educational institutions participating in Title IX financial aid programs to provide incoming college students with primary prevention and awareness programs addressing sexual violence. The Campus SaVE Act mandates that these programs include a component on bystander intervention. Currently, there is no comparable legislation regarding sexual assault among adolescents (e.g., mandating bystander programs in secondary schools).

Findings from this review indicate that bystander programs have beneficial effects on bystander intervention behavior. This review therefore provides important evidence of the effectiveness of mandated programs on college campuses. Additionally, results from the moderator analyses indicated that the effects on bystander intervention are similar for adolescents and college students, which suggests that early implementation of bystander programs (i.e., in secondary schools with adolescents) may be warranted.

Importantly, although we found that bystander programs had a significant effect on bystander intervention behavior, there was no evidence that these programs had an effect on participants’ sexual assault perpetration. Although most bystander sexual assault prevention programs aim to shift attitudes in the hopes of preventing sexual assault perpetration, this review provided no evidence that these programs decrease participants’ perpetration rates. This suggests that bystander programs may be appropriate for targeting bystander behavior but may not be appropriate for targeting the behavior of potential perpetrators. Additionally, effects of bystander programs on bystander intervention diminished six-months post-intervention. Thus, booster sessions may be needed to yield any sustained intervention effects.

Implications for research

Findings from this review suggest there is a fairly strong body of research assessing the effects of bystander programs on attitudes and behaviors. However, there are important questions worth further exploration.

First, according to one prominent logic model, bystander programs promote bystander intervention by fostering prerequisite knowledge and attitudes (Burn, 2009). Our meta-analysis provides inconsistent evidence of the effects of bystander programs on knowledge and attitudes, but promising evidence of short-term effects on bystander intervention behavior. These results cast uncertainty on the proposed relationship between knowledge/attitudes and bystander behavior. However, our methods do not permit any formal evaluation of this relationship. The field’s understanding of the causal mechanisms of program effects on bystander behavior would benefit from further analysis (e.g., path analysis mapping relationships between specific knowledge/attitude effects and bystander intervention).
Second, bystander programs exhibit a great deal of content variability, most notably in framing sexual assault as a gendered or gender-neutral problem. That is, bystander programs tend to adopt one of two main approaches to addressing sexual assault: (1) they present sexual assault as a gendered problem (overwhelmingly affecting women) or (2) they present sexual assault as a gender-neutral problem (affecting women and men alike). Differential effects of these two types of programs remain largely unexamined. Our analysis indicated that (1) the sex of victims/perpetrators presented in interventions (i.e., portrayed in programs as gender neutral or male perpetrator and female victim) and (2) whether programs were implemented in mixed- or single-sex settings were not significant predictors of program effects on bystander intervention. However, these findings are limited to a single outcome and they should be considered preliminary, as they are based on a small sample ($n = 11$). The field’s understanding of the differential effects of gendered versus gender neutral programs would benefit from the design and implementation of high-quality primary studies that make direct comparisons between these two types of programs (e.g., RCTs comparing the effects of two active treatment arms that differ in their gendered approach).

Finally, as previously noted, all but two eligible studies were conducted in the United States. Thus, high-quality studies conducted outside of the United States are needed to provide a global perspective on the efficacy of bystander programs.
Background

The problem, condition or issue

Sexual assault among adolescents and college students

Sexual assault is a significant problem among adolescents and college students in the United States and globally. Findings from the Campus Sexual Assault study estimated that 15.9% of American college women had experienced attempted or completed sexual assault (i.e., unwanted sexual contact that could include sexual touching, oral sex, intercourse, anal sex, or penetration with a finger or object) prior to entering college and 19% had experienced attempted or completed sexual assault since entering college (Krebs, Lindquist, Warner, Fisher, & Martin, 2009).

Similar rates have been reported in Australia (Australian Human Rights Commission, 2017), Chile (Lehrer, Lehrer, & Koss, 2013), China (Su, Hao, Huang, Xiao, & Tao, 2011), Finland (Bjorklund, Hukkanen-Nyholm, Huttunen, & Kunttu, 2010), Poland (Tomaszewska & Krahe, 2015), Rwanda (Van DeCraen, Michielsen, Herbots, Rossem, & Temmerman, 2012), Spain (Vazquez, Torres, Otero, 2012) and in a global survey of countries in Africa, Asia, and the Americas (Pengpid & Peltzer, 2016).

These rates are problematic, as sexual assault in adolescence and/or young adulthood is associated with numerous adverse outcomes, including risk of repeated victimization, depressive symptomology, heavy drinking, and suicidal ideation (Cortens, Eckenrode, & Rothman, 2013; Cui, Ueno, Gordon, & Fincham, 2013; Halpern, Spriggs, Martin, & Kupper, 2009). Importantly, there is evidence indicating experiences of sexual assault during these two life phases are related, as victimization and perpetration during adolescence are, respectively, associated with increased risk of victimization and perpetration during young adulthood (Cui, Ueno, Gordon, & Fincham, 2013). Thus, early prevention efforts are of paramount importance.

Reviews of research on the effectiveness of programs designed to prevent sexual assault among adolescents and college students have noted both a dearth of high-quality studies, such as randomized controlled trials (RCTs), and minimal evidence that these prevention programs have meaningful effects on young people’s behavior (DeGue, Valle, Holt, Massetti, Matjasko, & Tharp, 2014; De Koker, Mathews, Zuch, Bastien, & Mason-Jones, 2014). Concerning the latter point, evaluations of such programs tend to measure attitudinal outcomes (e.g., rape supportive attitudes, rape myth acceptance) more frequently than behavioral outcomes (e.g., perpetration or victimization) (Anderson & Whiston, 2005; Cornelius & Resseguie, 2007; DeGue et al., 2014).
Additionally, findings from a meta-analysis of studies assessing outcomes of college sexual assault prevention programs suggested that effects are larger for attitudinal outcomes than for the actual incidence of sexual assault (Anderson & Whiston, 2005).

The intervention

The bystander approach

Given this paucity of evidence regarding behavior change, it is imperative to identify effective strategies for preventing sexual assault among adolescents and young adults. One promising strategy is the implementation of bystander programs, which encourage young people to intervene when witnessing incidents or warning signs of sexual assault (e.g., controlling behavior, such as intervening with a would-be perpetrator leading an intoxicated person into an isolated area). The strength of the bystander model lies in its emphasis on the role of peers in the prevention of violence. Peers are a salient influence on young people's intimate relationships (Adelman & Kil, 2007; Giordano, 2003). In some respects, this influence can be detrimental, as having friends involved in violent intimate relationships (i.e., characterized by sexual or physical violence) is a risk factor for becoming both a perpetrator and victim of violence (Arriaga & Foshee, 2004; Foshee, Benefield, Ennett, Bauman, & Suchindran, 2004; Foshee, Linder, MacDougall, & Bangdiwala, 2001; Foshee, Reyes, & Ennett, 2010; McCauley et al., 2013). However, peers can also have a positive impact on intimate relationships.

Young victims and perpetrators of violence are often reluctant to divulge their experience or to seek help (especially from adults), but when they do seek help they often seek it from their peers (Ashley & Foshee, 2005; Black, Tolman, Callahan, Saunders, & Weisz, 2008; Molidor & Tolman, 1998; Weisz, Tolman, Callahan, Saunders, & Black, 2007). Victims may trust their peers to provide a valuable source of support after an assault has occurred, and as such, peers have the potential to play a pivotal role in the prevention of sexual assault by intervening when they witness its warning signs. In fact, in a contemporary “hookup culture” adolescents and young adults are more likely to meet and socialize in groups than they are to date in pairs and, therefore, warning signs of assault are frequently exhibited in communal spaces (Bogle, 2007; 2008; England & Ronen, 2015; Molidor & Tolman, 1998; Wade, 2017). Thus, the social nature of intimate relationships during these life stages can make peers pivotal actors in the prevention of sexual assault.

However, the potential for peer intervention can be undermined by a general “bystander effect” that diffuses responsibility for action in group settings (Darley & Latane, 1968). To intervene as a witness to sexual assault, individuals must notice the event (or its warning signs), define the event as warranting action/intervention, take responsibility for acting (i.e., feel a sense of personal duty), and demonstrate a sufficient level of self-efficacy (i.e., perceived competence to successfully intervene) (Latane & Darley, 1969). Studies have indicated that, as witnesses to sexual assault, young people often fail to meet these criteria (Banyard, 2008; Bennett, Banyard, & Garnhart, 2014; Burn, 2009; Casey & Ohler, 2012; Exner & Cummings, 2011; McCauley et al., 2013; McMahon, 2010; Noonan & Charles, 2009), with males being less likely than females to intervene (Banyard, 2008; Burn, 2009; Edwards, Rodenhizer-Stampfl, & Eckstein, 2015; Exner & Cummings, 2011; McMahon, 2010).
Thus, bystander programs seek to sensitize young people to warning signs of sexual assault, create attitudinal change that fosters bystander responsibility for intervening (e.g., creating empathy for victims), and build requisite skills/tactics for taking action (Banyard, 2011; Banyard, Plante, & Moynihan, 2004; Burn, 2009; McMahon & Banyard, 2012). Many of these programs are implemented with large groups of adolescents or college students in the format of a single training/education session (e.g., as part of college orientation). However, some programs use broader implementation strategies, such as advertising campaigns where signs are posted across college campuses to encourage students to act when witnessing signs of violence. The bystander model was developed and popularized in the US but has been adapted for use in global contexts.

**How the intervention might work**

By treating young people as potential allies in preventing sexual assault, bystander programs have the potential to be less threatening than traditional sexual assault prevention programs, which tend to approach young people as either potential perpetrators or victims of sexual violence (Burn, 2009; Messner, 2015; [Jackson] Katz, 1995). Instead of placing emphasis on how young people may modify their individual behavior to either respect the sexual boundaries of others or reduce their personal risk for being sexually assaulted, bystander programs aim to foster prerequisite knowledge and skills for intervening on behalf of victims. Thus, by treating young people as part of the solution to sexual assault, rather than part of the problem, bystander programs limit the risk of defensiveness or backlash among participants (e.g., decreased empathy for victims, increased rape myth acceptance) (Banyard et al., 2004; Katz, 1995).

In addition to encouraging young people to prevent sexual violence through bystander intervention, these programs may reduce the likelihood of participants committing sexual assault themselves. To illustrate, Katz (1995) describes the rationale behind the Mentors in Violence Prevention (MVP) bystander program as follows: “rather than focus on men as actual or potential perpetrators [emphasis in original], we focus on them in their role as potential bystanders [emphasis in original]. This shift in emphasis greatly reduces the participants’ defensiveness” (p. 168). In other words, young men may be more receptive to prevention programming, and thus less likely to commit sexual assault, when they are approached as part of the solution rather than part of the problem to sexual assault. Consistent with this line of reasoning, studies on the effects of bystander programs often measure participants’ rates of sexual assault as a program outcome (Foubert, 2000; Foubert, Newberry, & Tatum, 2007; Gidycz, Orchowski, & Berkowitz, 2011; Miller et al., 2014; Miller et al., 2013; Salazar, Vivolo-Kantor, Hardin, & Berkowitz, 2014).

As outlined by Burn (2009) bystander programs are designed to promote the following prerequisites for intervention: noticing an event, identifying a situation as warranting intervention, taking responsibility for acting, and deciding how to help. This often involves educating young people about what constitutes sexual assault, portraying victims as worthy of assistance, and building skills necessary to intervene (e.g., providing strategies for what to do and say). Although most bystander programs share the common goal of promoting such prerequisites for intervention, their specific program content and framing of sexual assault varies.
Research has indicated that, relative to males, females are overwhelmingly the victims of sexual assault (Foshee, 1996; Gressard, Swahn, & Tharp, 2015; Harned, 2001; Howard, Wang, & Yan, 2007). Thus, the earliest bystander programs tended to apply a gendered perspective to the prevention of sexual assault among adolescents and college students. For example, Katz (1995) developed MVP with the goal of inspiring male college athletes to challenge sociocultural definitions of masculinity that equate men’s strength with dominance over women. At the time of its inception, MVP was unique in its explicit focus on masculinity as well as its nonthreatening bystander approach that encouraged young men to intervene when witnessing acts (or warning signs) of violence against women. As Katz explained, MVP reduces young men’s defensiveness to violence prevention efforts by focusing on men as potential bystanders to violence, rather than potential perpetrators of violence. In addition to reducing men’s defensiveness to intervention efforts, this bystander approach emphasizes the point that “when men don’t speak up or take action in the face of other men’s abusive behavior toward women, that constitutes implicit consent of such behavior” (Katz, 1995, p. 168).

Since the inception of MVP a number of programs have emerged to address barriers to bystander intervention among adolescents and college students. Although they all share the common goal of inspiring bystanders to act in ways that prevent sexual assault, these programs exhibit a great deal of variation in scope pertaining to their target bystander populations (i.e., males and/or females, secondary school or college students), sex of victims, and gendered versus gender-neutral approach. For example, some programs use a gendered approach by (1) critiquing gender norms that can promote violence against women and (2) encouraging males to intervene on behalf of female victims (e.g., MVP, see Katz, 1995). Others use a gender-neutral approach to build a sense of community responsibility to intervene on behalf of both male and female victims of sexual assault (e.g., Bringing in the Bystander, see Banyard, Moynihan, & Crossman, 2009; Banyard, Moynihan, & Plante, 2007). One of the major differences between gendered and gender-neutral bystander programs is that the former places socio-cultural forces, such as gender norms, at the center of discussions of violence whereas the latter places the bystander, and individual cognitive processes when encountering violence, at the center of discussions of violence ([Jackson] Katz, Hesiterkamp, & Fleming, 2011; Messner, 2015).

Comparing the effects of gendered and gender-neutral programs has the potential to identify important determinants of the success of bystander programs. Although there is no empirical examination of the different effects of these programs, there are theoretical reasons to believe that each has the potential to be successful under certain conditions. Namely, gendered approaches to bystander education programs may be better suited to target socio-cultural facilitators of sexual assault against women and address different patterns of bystander behaviors exhibited by males and females (Banyard, 2008; Burn, 2009; Exner & Cummings, 2011; Katz et al., 2011; [Jennifer] Katz, 2015; [Jennifer] Katz, Colbert, & Colangelo, 2015; McCauley et al., 2013; McMahon, 2010; Messner, 2015). On the other hand, gender-neutral programs may have the benefit of deflecting the criticism that prevention programs utilizing a gendered approach are inherently anti-male (Katz et al., 2011; Messner, 2015). Proponents of such programs assert that avoidance of such criticism is paramount to the success of sexual assault prevention programs. In their view, adolescents and young adults who are coming of age in a “post-feminist” era may be likely to reject gendered explanations of sexual assault and, instead, may respond more positively to gender-neutral
programs that use inclusive language that can be applied to a broad range of victims and perpetrators (Barreto & Ellemers, 2005; Kettrey, 2016; Swim, Aikin, Hall, & Hunter, 1995).

**Why it is important to do the review**

Policymakers in the United States perceive bystander programs to be beneficial, as evidenced by the 2013 Campus Sexual Violence Elimination (SaVE) Act’s requirement that post-secondary educational institutions participating in Title IX financial aid programs provide incoming college students with primary prevention and awareness programs addressing sexual violence. The Campus SaVE Act mandates that these programs include a component on bystander intervention. Currently, there is no comparable legislation regarding sexual assault among adolescents (e.g., mandating bystander programs in secondary schools). This is an unfortunate oversight, as adolescents who experience sexual assault are at an increased risk of repeated victimization in young adulthood (Cui et al., 2013). Thus, the implementation of bystander programs in secondary schools not only has the potential to reduce sexual assault among adolescents but may also have the long-term potential to reduce sexual assault on college campuses.

Findings from this systematic review will provide valuable evidence of the extent to which bystander programs, as mandated by the Campus SaVE Act, are effective in preventing sexual assault among college students. Additionally, by examining effects of these programs among adolescents, this review will provide educators and policy makers with information for determining whether such programs should be widely implemented in secondary schools. Currently, there are no Campbell or Cochrane Collaboration Reviews evaluating the effects of bystander programs on sexual assault among adolescents and/or college students. Of modest relevance to the proposed review, the Campbell and Cochrane Collaboration libraries include meta-analyses of the effects of more general programs (not bystander programs) designed to prevent or reduce relationship/dating violence among adolescents and/or young adults (De La Rue, Polanin, Espelage, & Pigott, 2014; Fellmeth, Heffernan, Nurse, Habibula, & Sethi, 2013). Both of these reviews reported violence outcomes as aggregate measures that do not distinguish sexual violence from other forms of violence. Although they each found some evidence of significant effects on knowledge or attitudes pertinent to violence, neither found evidence of significant effects on young people’s behavior (i.e., rates of perpetration or victimization).

Two reviews published outside of the Campbell and Cochrane Collaboration libraries are of closer relevance to this review. These include a meta-analysis of the effects of bystander programs on sexual assault on college campuses (Katz & Moore, 2013) and a narrative review of studies examining the effects of bystander programs on dating and sexual violence among adolescents and young adults (Storer, Casey, & Herrenkohl, 2015).

In what they called an “initial” meta-analysis of experimental and quasi-experimental studies published through 2011, Katz & Moore (2013) found moderate beneficial effects of bystander programs on participants’ self-efficacy and intentions to intervene, and small (but significant) effects on bystander behavior, rape-supportive attitudes, and rape proclivity (but not perpetration). Effects were generally stronger among younger samples and samples containing a higher
percentage of males. The stronger effect for younger participants (i.e., younger college students) suggests such programs may be particularly effective with adolescents.

In a narrative review of studies examining the effects of bystander programs on dating violence and sexual assault among adolescents and young adults, Storer et al. (2015) highlighted beneficial effects on bystander self-efficacy and intentions but noted less evidence of beneficial effects on actual bystander behavior or perpetration of violence. While informative, each of these reviews has limitations. Katz & Moore’s (2013) meta-analysis focused exclusively on sexual assault on college campuses and did not examine effects of such programs among adolescents. Although Storer et al. (2015) focused on studies examining violence among both adolescents and young adults, their sample was limited in that it was exclusively comprised of peer-reviewed articles (i.e., the sample explicitly excluded theses, dissertations, and other gray literature). Additionally, the authors specified no research design criteria for inclusion (i.e., the sample included low-quality studies such as those utilizing single group pre- and post-test designs), limiting the strength of their conclusions. Importantly, Storer et al. reported no meta-analytic findings. Thus, to our knowledge there are currently no existing meta-analyses examining the effects of bystander programs on attitudes and behaviors regarding sexual assault among both college students and adolescents. Additionally, Katz & Moore’s early meta-analysis only included studies published/reported through 2011 (two years prior to the 2013 Campus SaVE Act) and did not evaluate program content as a moderator.

Our review examined the effects of bystander programs on attitudes (i.e., perceptions of violence/victims, self-efficacy to intervene, intentions to intervene) and behaviors (i.e., actual intervention behavior, perpetration) regarding sexual assault among adolescents and college students. Importantly, we present meta-analytic findings to quantitatively assess the influence of moderators (e.g., gender composition of sample, mean age of sample, education level of sample, single- or mixed- sex implementation, gendered content of program, fraternity/sorority membership, athletic team membership) on the effects of bystander programs.
Objectives

The problem, condition or issue

The overall objective of this systematic review and meta-analysis was to examine the effects bystander programs have on preventing sexual assault among adolescents and college students. More specifically, and given the study designs that were included, this review addressed three objectives.

1. The first objective was to assess the overall effects (including adverse effects), and the variability of the effects, of bystander programs on adolescents’ and college students’ attitudes and behaviors regarding sexual assault. This included general attitudes toward violence and victims, prerequisite skills and knowledge to intervene, self-efficacy to intervene, intentions/willingness to intervene when witnessing signs of sexual assault, actual intervention behavior, and perpetration of sexual assault.

2. The second objective was to explore the comparative effectiveness of bystander programs for different profiles of participants (e.g., mean age of the sample, education level of the sample, proportion of males/females in the sample, proportion of fraternity/sorority members in the sample, proportion of athletic team members in the sample).

3. The third objective was to explore the comparative effectiveness of different bystander programs in terms of gendered content and approach (e.g., conceptualizing sexual assault as a gendered or gender-neutral problem, mixed- or single-sex group implementation).
Methods

Criteria for considering studies for this review

Types of studies
To be eligible for inclusion in the review, studies must have used an experimental or controlled quasi-experimental research design to compare an intervention group (i.e., students assigned to a bystander program) with a comparison group (e.g., students not assigned to a bystander program). We limited our review to such study designs because these typically have lower risk of bias relative to other study designs (e.g., single group designs). More specifically, we included the following designs:

1. Randomized controlled trials: Studies in which individuals, classrooms, schools, or other groups were randomly assigned to intervention and comparison conditions.

2. Quasi-randomized controlled trials: Studies where assignment to conditions was quasi-random, for example, by birth date, date of week, student identification number, month, or some other alternation method.

3. Controlled quasi-experimental designs: Studies where participants were not assigned to conditions randomly or quasi-randomly (e.g., participants self-selected into groups). Given the potential selection biases inherent in these controlled quasi-experimental design, we only included those that also met one of the following criteria:

   a. Regression discontinuity designs: Studies that used a cutoff on a forcing variable to assign participants to intervention and comparison groups, and assessed program impacts around the cutoff of the forcing variable.

   b. Studies that used propensity score or other matching procedures to create a matched sample of participants in the intervention and comparison groups. To be eligible for inclusion, these studies must have also provided enough statistical information to permit estimation of pretest effect sizes for the matched groups.

   c. For studies where participants in the intervention and comparison groups were not matched, enough statistical information must have been reported to permit estimation of pretest effect sizes for at least one outcome measure.
Consistent with Campbell Collaboration policies and procedures, studies using experimental and quasi-experimental research designs were synthesized separately in the meta-analyses, given that experimental study designs have the highest level of internal validity. Furthermore, we collected extensive data on the risk of bias and study quality of all eligible studies, which we attended to when interpreting the findings from the systematic review and meta-analyses (described in greater detail below).

**Types of participants**

The review focused on studies that examined outcomes of bystander programs that target sexual assault and are implemented with adolescents and/or college students in educational settings. Eligible participants included adolescents enrolled in grades 7 through 12 and college students enrolled in any type of undergraduate postsecondary educational institution. Eligible participant populations included studies that reported on general samples of adolescents and/or college students as well as studies using specialized samples such as those primarily consisting of college athletes, fraternity/sorority members, and single-sex samples. Study samples primarily consisting of post-graduate students were ineligible for inclusion; the mean age of samples could be no less than 12 and no greater than 25 to be included in the review.

**Types of interventions**

Eligible intervention programs were those that approach participants as allies in preventing and/or alleviating sexual assault among adolescents and/or college students. Some part of the program had to focus on ways that cultivate willingness for a person to respond to others who are at risk for sexual assault. All delivery formats were eligible for inclusion (e.g., in-person training sessions, video programs, web-based training, advertising/poster campaigns). There were no intervention duration criteria for inclusion.

Studies that reported bystander outcomes but did not meet the aforementioned intervention inclusion criterion were not eligible for inclusion. Additionally, studies that assessed outcomes of programs that aimed to facilitate pro-social bystander behavior, but that did not explicitly include a component addressing sexual assault (e.g., programs to prevent bullying) were not eligible for inclusion.

Eligible comparison groups must have received no intervention services targeting bystander attitudes/behavior or sexual assault. Thus, treatment-treatment studies that compared outcomes of individuals assigned to a bystander program versus those assigned to a general sexual assault prevention program were not eligible for inclusion. Eligible comparison groups may have received a sham or attention treatment expected to have no effect on bystander outcomes or attitudes/behaviors regarding sexual assault.

**Types of outcome measures**

We included studies that measured the effects of bystander programs on at least one of the following primary outcome domains:
1. General attitudes toward sexual assault and victims (e.g., victim empathy, rape myth acceptance).

2. Prerequisite skills and knowledge for bystander intervention as defined by Burn (2009) (e.g., noticing sexual assault or its warning signs, identifying a situation as appropriate for intervention, taking responsibility for acting/intervening, knowing strategies for helping/intervening).

3. Self-efficacy with regard to bystander intervention (e.g., respondents’ confidence in their ability to intervene).

4. Intentions to intervene when witnessing instances or warning signs of sexual assault.

5. Actual intervention behavior when witnessing instances or warning signs of sexual assault.

6. Perpetration of sexual assault (i.e., rates of perpetration among individuals assigned to the treatment or comparison group of a study).

Depending on directionality, these outcomes capture both beneficial and adverse effects (e.g., increases or decreases in victim empathy, pro-social bystander behavior, etc.) that are important to adolescents, college students, and decision-makers alike.

Any outcome falling in these domains was eligible for inclusion. This includes outcomes measured with any form of assessment (e.g., self-report, official/administrative report, observation, etc.) that could be summarized by any type of quantitative score (e.g., percentage, continuous variable, count variable, categorical variable, etc.). In the event that a particular study included multiple measures of a single construct category (e.g., two measures of rape myth acceptance or bystander intervention within a given study), we only included one outcome per study for that construct. We selected the most similar outcomes for synthesis within a construct category (described in detail below).

**Duration of follow-up**

Studies reporting follow-ups of any duration were eligible for inclusion. When studies reported more than one follow-up wave, each wave was coded and identified by its reported duration. As described in more detail below, follow-ups of similar durations were analyzed together.

**Types of settings**

The review focused on studies that examine outcomes of bystander programs that target sexual assault and are implemented with adolescents and/or college students in educational settings. Eligible educational settings included secondary schools (i.e., grades 7-12) and colleges or universities. Studies that assessed bystander programs implemented with adolescents and young adults outside of educational institutions (e.g., in community settings, military settings) were ineligible for inclusion in the review. There were no geographic limitations on inclusion criteria. Research conducted in any country was eligible.
Search methods for identification of studies

Search strategy
We identified candidate studies through searches of electronic databases, relevant academic journals, and gray literature sources. We also contacted leading authors and experts on bystander programs to identify any current/ongoing research that might be eligible for the review. Additionally, we screened the bibliographies of eligible studies and relevant reviews to identify additional candidate studies. We conducted forward citation searches (searches for reports citing eligible studies) using the website Google Scholar, as this database produces similar results to other search engines (e.g., Web of Science; Tanner-Smith & Polanin, 2015) and is also more likely to locate existing gray literature. Our search was global in scope and attempted to identify studies of bystander programs implemented in any country.

Electronic searches
The prevention of sexual assault among college students and adolescents is a topic that spans multiple disciplines (e.g., sociology, psychology, education, public health). Thus, we searched a variety of databases that are relevant to these fields. Search terms varied by database, but generally included two blocks of terms and appropriate Boolean or proximity operators, when allowed: blocks included terms that address the intervention and outcomes. We specifically searched the following electronic databases (hosts) in October 2016 and June 2017:

- Cochrane Central Register of Controlled Trials (CENTRAL)
- Cochrane Database of Abstracts of Reviews of Effects (DARE)
- Education Resources Information Center (ERIC, via ProQuest)
- Education Database (via ProQuest)
- International Bibliography of the Social Sciences (IBSS, via ProQuest)
- PsycINFO (via ProQuest)
- PsycARTICLES (via ProQuest)
- PubMed
- Social Services Abstracts (via ProQuest)
- Sociological Abstracts (via ProQuest)

The strategy for searching electronic databases involved the use of search terms specific to the types of interventions and outcomes eligible for inclusion. Search terms for types of interventions included general terms for bystander programs as well as names of specific bystander programs (e.g., Mentors in Violence Prevention). Search terms for types of outcomes included terms that are specific to measures of sexual violence (e.g., sexual assault) as well as more general terms that have the potential to identify studies that measure physical and/or sexual violence. Due to the overwhelming focus of bystander programs on adolescents and college students (aside from a few implementations with military samples) search terms did not limit initial results by the age or general characteristics of the target population.
The search terms and strategy for PsycINFO via ProQuest were as follows (terms were modified for other databases):

(AB, TI(“bystander”)) AND (AB, TI(“education” OR “program” OR “training” OR “intervention” OR “behavior” OR “attitude” OR “intention” OR “efficacy” OR “prosocial” OR “pro-social” OR “empowered” OR “Bringing in the Bystander” OR “Green Dot” OR “Step Up” OR “Mentors in Violence Prevention” OR “MVP” OR “Know Your Power” OR “Hollaback” OR “Circle of 6” OR “That’s Not Cool” OR “Red Flag Campaign” OR “Where Do You Stand” OR “White Ribbon Campaign” OR “Men Can Stop Rape” OR “The Men’s Program” OR “The Women’s Program” OR “The Men’s Project” OR “Coaching Boys into Men” OR “Campus Violence Prevention Program” OR “Real Men Respect” OR “Speak Up Speak Out” OR “How to Help a Sexual Assault Survivor” OR “InterACT”)) AND (AB, TI(“sexual assault” OR “rape” OR “violence” OR “victimization”))

**Searching other resources**

We searched the tables of contents of current issues of journals that publish research on sexual violence. This included the following journals: *Journal of Adolescent Health, Journal of Community Psychology, Journal of Family Violence, Journal of Interpersonal Violence, Psychology of Violence, Violence Against Women,* and *Violence and Victims.* We searched these sources in October 2016 and June 2017.

We also conducted gray literature searches to identify unpublished studies that met inclusion criteria. This included searching electronic databases that catalog dissertations and theses, searching conference proceedings, and searching websites with content relevant to sexual assault and/or violence against women. We specifically searched the following gray literature sources in October 2016 and June 2017:

- ProQuest Dissertations and Theses Global
- Clinical Trials Register (https://clinicaltrials.gov)
- End Violence Against Women International - conference proceedings (http://www.evawintl.org/conferences.aspx)
- National Sexual Violence Resource Center website (nsvrc.org)
- National Online Resource Center on Violence Against Women website (VAWnet.org)
- Us Department of Justice Office on Violence Against Women website (www.justice.gov/ovw)
- Center for Changing our Campus Culture website (www.changingourcampusculture.org)

Additionally, we searched reference lists of previous systematic reviews and meta-analyses, CVs and websites of primary authors of eligible studies, and reference lists of eligible studies. We also conducted forward citation searches of all eligible studies.
Data collection and analysis

Selection of studies

Once candidate studies were identified in the literature search, each reference was entered into the project database as a separate record. Two reviewers then independently screened each study title and abstract and recorded their eligibility recommendation (i.e., ineligible or eligible for full-text screening) into the pertinent database record. Disagreements between reviewers were resolved by discussion and consensus, and the final abstract screening decision was recorded in the database. Potentially eligible studies were then retrieved in full text and these full texts were reviewed for eligibility, again using two independent reviewers who recorded their eligibility recommendation (and, when applicable, rationale for an ineligibility recommendation). Disagreements between reviewers were again resolved via discussion and consensus and the final eligibility decision was recorded in the database. In cases where we could not determine eligibility due to missing information in a report, we contacted study authors for this information.

Throughout the search and screening process we maintained a document that included the number of unique candidate studies identified through various sources (e.g., electronic database searches, academic journal searches, and gray literature searches). We used the information in this record to create a PRISMA flow chart that reports the screening process (Moher et al., 2015). Additionally, we used the final screening decisions recorded in the meta-analysis database to create a table that lists studies excluded during the full-text screening phase along with the rationale for each exclusion decision.

Data extraction and management

Two reviewers independently double-coded all included studies, using a piloted codebook (see Appendix). All coding was entered into an electronic database, with a separate record maintained for each independent coding of each study. Coding disagreements were resolved via discussion and consensus with final coding decisions maintained in a separate record. If data needed to calculate an effect size were missing from a report, we contacted the primary study authors for this information.

The primary categories for coding were as follows: participant demographics and characteristics (e.g., age, gender, education level, race/ethnicity, athletic team membership, fraternity/sorority membership); intervention setting (e.g., state, country, secondary or post-secondary institution, mixed- or single-sex group); study characteristics (e.g., attrition, duration of follow-up, study design, participant dose, sample N); outcome construct (e.g., type, description of measure); and outcome results (e.g., timing at measurement, baseline and follow-up means and standard deviations or proportions).

Assessment of risk of bias in included studies

We assessed risk of bias in included studies using the Cochrane risk of bias tools for randomized studies (Higgins & Green, 2011) and non-randomized studies (Sterne et al., 2016). The Cochrane risk of bias tool for randomized studies assesses risk of bias in the following domains: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of
outcome assessment, incomplete outcome data, selective reporting, and other bias. These domains are each rated as low, unclear, or high risk of bias.

The Cochrane risk of bias tool for non-randomized studies, ROBINS-I, requires that at the protocol stage of the review, two sets of items are determined: (1) confounding areas that are expected to be relevant to all or most studies in the review and (2) co-interventions that could be different between intervention groups with the potential to differentially impact outcomes. In our protocol, we anticipated the following confounding factors, which were coded in the risk of bias assessments for non-randomized studies: gender, fraternity/sorority membership, athletic team membership, pre-intervention attitudes (e.g., victim empathy, rape myth acceptance), pre-intervention bystander measures (e.g., efficacy, intentions, behavior), and prior sexual assault victimization. Additionally, we anticipated the following co-interventions to have a potential impact on outcomes, and they were coded in the risk of bias assessments: general sexual assault prevention programs, dating violence prevention programs, and general bystander programs (not explicitly targeting sexual violence).

Measures of treatment effect
We extracted relevant summary statistics (e.g., means and standard deviations, proportions, observed sample sizes) to calculate effect sizes. We then used David Wilson’s (2013) online effect size calculator to calculate effect sizes. The overwhelming majority of studies reported continuous measures of treatment effects, so we used a standardized mean difference (SMD) effect size metric with a small sample correction (i.e., Hedges’ $g$). In the rare cases in which binary outcome measures were reported in included studies, we deemed these measures to represent the same underlying construct as continuous measures, as they typically relied on the same measurement tools as relevant continuous measures. Thus, we transformed any log odd ratio effect sizes available from binary measures into standardized mean difference effect sizes by entering the observed proportions and sample sizes into Wilson’s (2013) online effect size calculator. All standardized mean difference effect sizes were coded such that positive values (i.e., greater than 0) indicate a beneficial outcome for the intervention group.

Unit of analysis issues
The unit of analysis of interest for this review was the individual (i.e., individual-level attitudes and behaviors). Nine of the included studies used cluster randomized trial designs where participants were randomized into the intervention or comparison conditions at the group level (e.g., entire schools were assigned to a single condition), and inferences were made at the individual level. To correct for these unit of analysis errors, we followed the procedures outlined in the Cochrane Handbook (Higgins & Green, 2011) to inflate the standard errors of the effect sizes from these nine studies by multiplying them by the design effect:

$$1 + (M - 1) \text{ICC},$$

where $M$ is the average cluster size for a given study and ICC is the intracluster correlation coefficient for a given outcome. In cases where study authors did not report ICCs we used a liberal assumed value of .10. This value has been used in a past Campbell review that synthesizes attitude
and behavior outcomes of dating/sexual violence prevention programs (De La Rue et al., 2014) and is supported by Hedges and Hedberg’s (2007) research on ICCs in cluster randomized trials conducted in educational settings.

**Dealing with missing data**

When studies reported insufficient data to calculate effect sizes we contacted the primary authors to request the necessary information. This author query method was highly successful; there were only two cases in which we were unable to obtain sufficient data from study authors (see details below with search results).

**Assessment of heterogeneity**

We assessed and reported heterogeneity using the $\chi^2$ statistic and its corresponding $p$ value, as well as the $I^2$ and $\tau^2$ statistics. We used the restricted maximum likelihood estimator for $\tau^2$. When at least ten studies were included in any given meta-analysis, we also used mixed-effect meta-regression models to conduct moderator analyses. Moderators specified in the protocol included: (1) gendered content of the program; (2) mixed- or single-sex group implementation; (3) gender composition of the sample; (4) education level of the sample (i.e., secondary school or college students); (5) mean age of the sample; (6) proportion of fraternity/sorority members in the sample; (7) and proportion of athletic team members in the sample.

**Assessment of reporting biases**

To assess the potential of small study/publication bias, we used the contour enhanced funnel plot (Palmer, Peters, Sutton, & Moreno, 2008), Egger’s regression test (Egger, Smith, Schneider, & Minder, 1997), and trim and fill analysis (Duval & Tweedie, 2000).

**Data synthesis**

We conducted all statistical analyses with the metafor package in R. We conducted meta-analyses using random-effects inverse variance weights and reported 95% confidence intervals along with all mean effect size estimates. We conducted and reported all meta-analyses separately for RCTs and non-RCTs. Additionally, we conducted all syntheses separately by outcome domain (i.e., attitudes toward sexual assault and victims, prerequisite skills and knowledge to intervene, bystander self-efficacy, bystander intentions, bystander behavior, perpetration) and follow-up timing. For most outcomes, follow-up timing fell into the following categories: immediate post-intervention, one- to four-months post-intervention, and six months to one-year post-intervention. We displayed each synthesis using forest plots.

To minimize any potential bias in the meta-analysis results due to effect size outliers, prior to conducting the meta-analysis, we Winsorized any outlier effect sizes that fell more than two standard deviations away from the mean of the effect size distribution (Lipsey & Wilson, 2001). In such cases, we replaced the outlier effect size with the value that fell exactly two standard deviations from the mean of the distribution of effect sizes.
As noted previously, each meta-analysis synthesized a set of statistically independent effect sizes. To ensure the statistical independence of effect sizes synthesized within any given meta-analysis, we split all analyses by follow-up timing and outcome domain. In the event that a particular study included multiple measures within a single outcome domain (e.g., two measures of rape myth acceptance within a given study), we only included one outcome per study for that construct. We selected the most similar outcomes for synthesis within a construct category. We adopted this approach for synthesizing statistically independent effect sizes given the small number of included studies and effect sizes available for synthesis. Other approaches that can be used to synthesize dependent effect sizes, such as robust variance estimation, require larger sample sizes for efficient parameter estimation than the sample sizes we had available for analysis (Hedges, Tipton, & Johnson, 2010; Tanner-Smith & Tipton, 2014; Tipton, 2013).

Subgroup analysis and investigation of heterogeneity

We performed sub-group analyses based on study design, synthesizing effect sizes (1) for RCTs alone, (2), for non-RCTs alone, and (3) for RCTs and non-RCTS combined. For each of these three subgroup analyses we assessed and reported heterogeneity using the $\chi^2$ statistic and its corresponding $p$ value, $I^2$, and $\tau^2$. When at least ten studies were included in any given meta-analysis, we used mixed-effect meta-regression models to conduct moderator analyses.

Sensitivity analysis

When at least ten studies were included in a given meta-analysis we conducted sensitivity analyses to examine whether study-level attrition and high risk of bias (for each domain assessed with the risk of bias tools) were associated with effect size magnitude, using mixed-effects meta-regression models. Additionally, we conducted sensitivity analyses that removed any Winsorized effect sizes from the meta-analysis, to assess whether this method for handling outliers may have substantively altered the review findings.

Interpretation of findings

To provide substantive interpretations of statistically significant mean effect sizes, we transformed the average standardized mean difference effect size back into an unstandardized metric using commonly reported scales/measures in the respective outcome domains. Namely, we multiplied the average standardized mean difference effect size by the standard deviation of a scale/measure that was frequently used in our meta-analytic sample to yield an unstandardized difference in means. We aimed to select means and standard deviations from the most representative studies in our sample (e.g., RCTs and/or studies with larger sample sizes), but we recognize this is an imperfect process that requires generalization from a single study (for each significant outcome). We present these transformations in our Discussion section, at the conclusion of this report. Readers should be mindful of the fact that these are extrapolations intended to provide meaningful context to our findings, and that results are most accurately represented by the Hedges’ $g$ effect sizes.
Results

Description of studies

Results of the search

We conducted an initial literature search in October 2016 and an updated search in June 2017. Figure 1 outlines the flow of studies through the search and screening process. Through our initial and updated search we identified 797 reports. Of these reports, 738 were identified from searches of electronic databases, 19 from ClinicalTrials.gov, 1 from conference proceedings, 5 from website searches, 20 from reference lists of review articles, 1 from tables of contents searches, 3 from reference lists of eligible reports, 9 from CVs and websites of primary authors of eligible studies, and 1 from forward citation searching of eligible studies. After deleting duplicate reports and reports that were deemed ineligible through the abstract screening process 154 reports were deemed eligible for full-text screening. Three of these reports could not be located; thus, we screened a total of 151 full-text reports for eligibility.
Included studies

Twenty-seven independent studies summarized in 38 reports met inclusion criteria. One report (Jouriles et al., 2016) presented findings from two independent studies. We coded these two studies separately and identified them as Jouriles et al. (2016a) and Jouriles at al. (2016b). Table 1 summarizes aggregate characteristics of all eligible studies. Twelve studies utilized random assignment at the individual level, nine studies utilized random assignment at the group level, and six studies utilized non-random assignment but reported pre-test equivalence measures on at least one eligible outcome. More specific details of each individual study are summarized in Table A1 (see Figures and Tables Appendix). The majority of studies were conducted in the United States, with one study conducted in Canada and one conducted in India. This geographic pattern was primarily due to a paucity of research on bystander programs conducted outside of the United States and, to a lesser extent, to the failure of studies conducted outside the United States to meet the scope and methodological criteria for this review.
Table 1: Characteristics of Included Studies (N = 27)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study Design</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized - Individual</td>
<td>12</td>
<td>44.44</td>
</tr>
<tr>
<td>Randomized - Group</td>
<td>9</td>
<td>33.33</td>
</tr>
<tr>
<td>Non-Randomized</td>
<td>6</td>
<td>22.22</td>
</tr>
<tr>
<td><strong>Comparison Treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active/Sham Treatment</td>
<td>9</td>
<td>33.33</td>
</tr>
<tr>
<td>Inactive Control</td>
<td>18</td>
<td>66.67</td>
</tr>
<tr>
<td><strong>Peer Reviewed</strong></td>
<td>22</td>
<td>81.48</td>
</tr>
<tr>
<td><strong>Funded</strong></td>
<td>18</td>
<td>66.67</td>
</tr>
<tr>
<td><strong>Study Report Year</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>2</td>
<td>7.41</td>
</tr>
<tr>
<td>2016</td>
<td>6</td>
<td>22.22</td>
</tr>
<tr>
<td>2015</td>
<td>4</td>
<td>14.81</td>
</tr>
<tr>
<td>2014</td>
<td>4</td>
<td>14.81</td>
</tr>
<tr>
<td>2012</td>
<td>1</td>
<td>3.70</td>
</tr>
<tr>
<td>2011</td>
<td>2</td>
<td>7.41</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>3.70</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>3.70</td>
</tr>
<tr>
<td>2007</td>
<td>2</td>
<td>7.41</td>
</tr>
<tr>
<td>2000</td>
<td>1</td>
<td>3.70</td>
</tr>
<tr>
<td>1998</td>
<td>1</td>
<td>3.70</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>3.70</td>
</tr>
<tr>
<td>Not Reported</td>
<td>1</td>
<td>3.70</td>
</tr>
<tr>
<td><strong>Educational Setting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College/University</td>
<td>22</td>
<td>81.48</td>
</tr>
<tr>
<td>Secondary School</td>
<td>5</td>
<td>18.52</td>
</tr>
<tr>
<td><strong>Country</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>25</td>
<td>92.59</td>
</tr>
<tr>
<td>Canada</td>
<td>1</td>
<td>3.70</td>
</tr>
<tr>
<td>India</td>
<td>1</td>
<td>3.70</td>
</tr>
</tbody>
</table>

Three studies were eligible for the review by methodological standards but did not report codable outcomes (Cook-Craig et al., 2017; Darlington, 2014; Mabry & Turner, 2016). For example, in Cook-Craig et al.’s (2017) study the unit of analysis was the school, not the individual. The study authors surveyed entire intervention and comparison schools across several years. As a result, the individual students composing the sample changed from wave to wave – and many students entering the school in later waves did not receive direct intervention. Darlington (2014) only reported within-group pre-post effect sizes. No between-group data were reported and our attempt to obtain these data from the study author was unsuccessful. Mabry and Turner (2016) reported three-way ANOVA findings from a three-armed study (i.e., one eligible intervention group, one ineligible intervention group, and a comparison group). Our attempts to obtain data for the eligible intervention and comparison groups from the study author were unsuccessful. We therefore
summarized each of these three studies in the systematic review but were unable to include results from those studies in any of the meta-analyses.

Two studies (Banyard et al., 2007; Jouriles et al., n.d.) included two eligible intervention arms. Banyard et al. (2007) randomly assigned participants to one of three groups: (1) a no-treatment control that only completed pre- and post-intervention surveys, (2) an intervention group that was assigned to complete one 90-minute bystander program session, or (3) an intervention group that was assigned to complete three 90-minute bystander program sessions. Jouriles et al. (n.d.) randomly assigned participants to one of three groups: (1) a comparison “sham treatment” condition that presented material on study skills, (2) a computer-delivered bystander program that participants completed independently, or (3) a computer-delivered bystander program that participants completed in a lab under supervision. We handled these multiple-arm studies by selecting the intervention groups that were most consistent with the intervention groups from the other studies in the sample (i.e., the single-session treatment for Banyard et al. and the unmonitored treatment for Jouriles et al.). We contrasted these intervention groups with their comparison groups and included the calculated effect sizes in our main meta-analyses. We then conducted sensitivity analyses in which we ran additional meta-analyses that included effect sizes that compared the alternative intervention arms with the comparison arm.

Excluded studies
As shown in Figure 1, 105 reports were deemed ineligible after full-text screening. These reports were ineligible because they either did not present an evaluation of an eligible intervention ($n = 25$), did not include an eligible comparison group ($n = 58$), did not involve an eligible population or setting ($n = 11$), did not involve an eligible research design ($n = 8$), or did not measure any eligible outcomes ($n = 3$). Table A2 provides a brief description of each of these studies as well as the specific reasons for exclusion (see Figures and Tables Appendix).

Risk of bias in included studies

Randomized studies
Table 2 summarizes the risk of bias for the 21 randomized studies included in the systematic review. A large percentage of studies failed to report information that would permit assessment of risk of bias; these were coded as unclear risk. Among those studies reporting sufficient information, most exhibited low risk of bias in the domains of random sequence generation, allocation concealment, blinding of participants, and selective reporting. The majority of randomized studies (95.2%) exhibited high risk of bias in blinding of outcome assessment; these studies relied on self-report outcome data. Although a slight majority of randomized studies (52.4%) exhibited low risk of bias in handling incomplete outcome data, one-third exhibited high risk of bias. These studies with high risk of bias tended to exhibit uneven attrition between the intervention and comparisons groups or handled missing data inappropriately (e.g., mean imputation). The majority of randomized studies (81.0%) exhibited high risk of bias in the “other” domain; these studies tended to be conducted by researchers who either developed the
intervention under evaluation or were closely affiliated with the developer of the intervention under evaluation.

**Table 2: Risk of Bias of Randomized Studies**

<table>
<thead>
<tr>
<th>Source of Bias</th>
<th>Low Risk</th>
<th></th>
<th></th>
<th>High Risk</th>
<th></th>
<th></th>
<th>Unclear Risk</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>Random Sequence Generation</td>
<td>8</td>
<td>38.1</td>
<td></td>
<td>13</td>
<td>61.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allocation Concealment</td>
<td>3</td>
<td>14.3</td>
<td></td>
<td>18</td>
<td>85.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Participants</td>
<td>1</td>
<td>4.8</td>
<td></td>
<td>20</td>
<td>95.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Outcome Assessment</td>
<td>1</td>
<td>4.8</td>
<td>20</td>
<td>95.2</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incomplete Outcome Data</td>
<td>11</td>
<td>52.4</td>
<td></td>
<td>7</td>
<td>33.3</td>
<td>3</td>
<td>14.3</td>
<td></td>
</tr>
<tr>
<td>Selective Reporting</td>
<td>19</td>
<td>90.5</td>
<td>2</td>
<td>9.5</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Bias</td>
<td>3</td>
<td>14.3</td>
<td>17</td>
<td>81.0</td>
<td>1</td>
<td>4.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(N = 21\) studies.

**Non-randomized studies**

Table 3 summarizes the risk of bias for the 6 non-randomized studies included in the systematic review. Among the studies that reported group differences in confounding variables, most reported congruence between groups. This was often a product of the targeted population. For example, three studies targeted a specific gender (i.e., young men or young women) and two specifically targeted fraternity or sorority members. Only one study reported group differences between participants’ previous experiences with co-interventions. The remaining studies either failed to report this information \(n = 3\) or reported it at the aggregate level \(n = 2\).

**Table 3: Risk of Bias of Non-Randomized Studies**

<table>
<thead>
<tr>
<th>Source of Bias</th>
<th>Even</th>
<th></th>
<th></th>
<th>Uneven Between</th>
<th></th>
<th></th>
<th>Not</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>Confounding Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>4</td>
<td>66.7</td>
<td>1</td>
<td>16.7</td>
<td>1</td>
<td>16.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraternity/Sorority</td>
<td>2</td>
<td>33.3</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>66.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Athlete</td>
<td>1</td>
<td>16.7</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>83.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-intervention Attitudes</td>
<td>3</td>
<td>50.0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>50.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Victimization</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16.7</td>
<td>5</td>
<td>83.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confounding Interventions</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16.7</td>
<td>5</td>
<td>83.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(N = 6\) studies.

**Synthesis of results**

**Victim empathy**

Two studies measured victim empathy as an intervention outcome (Brokenshire, 2015; Salazar et al., 2014). Both studies operationalized victim empathy using the Rape Empathy Scale (Deitz,
Blackwell, Daley, & Bentley, 1982), which asks participants to rate their agreement with statements such as “In general, I feel that rape is an act that is not provoked by the rape victim.” Both of these studies involved random assignment of participants to conditions at the individual level. Salazar et al. was published in a peer-reviewed journal and Brokenshire was an unpublished master’s thesis. Because the two studies measured the outcome at very different time points, we could not synthesize effect sizes. Brokenshire measured victim empathy as an immediate post-test outcome \( g = -0.44, 95\% \text{ CI} [-0.79, -0.09] \) and Salazar et al. reported victim empathy six months post-intervention \( g = 0.68, 95\% \text{ CI} [0.41, 0.95] \). Notably, effect sizes from these two studies are divergent, with Brokenshire indicating a significant negative intervention effect at immediate post-test and Salazar et al. indicating a significant positive effect six months post intervention.

**Rape myth acceptance**

Twelve studies measured rape myth acceptance as an intervention outcome. These studies most often operationalized rape myth acceptance using the Illinois Rape Myth Acceptance Scale (Payne, Lonsway, & Fitzgerald, 1998), which asks respondents to rate their agreement with statements such as, “rape happens when a man’s sex drive gets out of control.” All but four of these studies (Amar et al., 2015; Baker et al., 2014; Foubert & Marriott, 1997; Peterson et al., 2016) involved random assignment of participants to conditions. All but one study (Brokenshire, 2015) was published in a peer-reviewed outlet; this study was an unpublished master’s thesis. We collapsed bystander efficacy effect sizes into three follow-up intervals: (1) immediate post-test (i.e., zero weeks to one week), (2) one month to four months post-intervention, and (3) six to seven months post-intervention.

**Immediate post-test effects.** Seven studies reported rape myth acceptance as an immediate post-test outcome. As shown in the forest plot in Figure 2, the intervention effect for the two non-randomized studies was significant and positive \( g = 0.65, 95\% \text{ CI} [0.23, 1.08] \) with minimal between-study heterogeneity \( X^2 = 1.07 \ [p = .30], I^2 = 6.75\%, \tau^2 = 0.01 \). The intervention effect for the five randomized studies was also significant and positive \( g = 0.30, 95\% \text{ CI} [0.03, 0.57] \), with significant between-study heterogeneity \( X^2 = 12.35 \ [p = .01], I^2 = 71.90\%, \tau^2 = 0.06 \). When we synthesized the findings from the randomized and non-randomized studies together, the average intervention effect was significant and positive \( g = 0.37, 95\% \text{ CI} [0.13, 0.61] \) with significant between-study heterogeneity \( X^2 = 15.73 \ [p = .02], I^2 = 66.36\%, \tau^2 = 0.06 \). Thus, at immediate post-intervention, the results indicate that bystander programs have a positive (beneficial) effect on rape myth acceptance. However, the small sample size in this meta-analysis \( n < 10 \) precluded ad-hoc analysis of moderators or small study/publication bias.
Figure 3: Forest Plot of Bystander Intervention Effects on Rape Myth Acceptance

Sensitivity analysis in which Banyard et al. (2007) one-session intervention was replaced by the three-session intervention did not substantively change these findings. As illustrated in Figure 3, the treatment effect for the five randomized studies (which included Banyard et al., 2007) was significant and positive ($g = 0.37$, 95% CI [0.07, 0.67]) with significant between-study heterogeneity ($X^2 = 14.04$ [p = .01], $I^2 = 76.48\%$, $\tau^2 = 0.08$). When synthesizing all studies together, the average intervention was significant and positive ($g = 0.43$, 95% CI [0.18, 0.68]) with significant between-study heterogeneity ($X^2 = 16.47$ [p = .01], $I^2 = 68.55\%$, $\tau^2 = 0.07$).
Six studies reported rape myth acceptance one to four months post-intervention. As depicted in the forest plot in Figure 4 the intervention effect for the three non-randomized studies was significant and positive (g = 0.63, 95% CI [0.27, 1.00]) with non-significant heterogeneity (X² = 1.02 [p = .60], I² = 0.00%, τ² = 0). The intervention effect for the three randomized studies was significant and positive (g = 0.25, 95% CI [0.04, 0.46]) with non-significant heterogeneity (X² = 1.73 [p = .42], I² = 0.00%, τ² = 0). Across all studies, the average intervention effect was significant and positive (g = 0.36, 95% CI [0.14, 0.59]) with non-significant heterogeneity (X² = 5.90 [p = .32], I² = 23.73%, τ² = 0.02). This average intervention effect indicates that, at one- to four-months post-intervention, bystander programs have a positive (beneficial) effect on rape myth acceptance. However, the small sample size in this meta-analysis (n < 10) precluded ad-hoc analysis of moderators or small study/publication bias.
Figure 4: Forest Plot of Bystander Intervention Effects on Rape Myth Acceptance – One to Four Months Post-Intervention

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Baker et al. (2014)</td>
<td>0.82 [-0.20, 1.84]</td>
</tr>
<tr>
<td>Foubert &amp; Marriott (1997)</td>
<td>0.28 [-0.50, 1.06]</td>
</tr>
<tr>
<td>Peterson et al. (2016)</td>
<td>0.71 [0.26, 1.16]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>0.63 [0.27, 1.00]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Banyard et al. (2007)</td>
<td>0.19 [-0.10, 0.48]</td>
</tr>
<tr>
<td>Gidycz et al. (2011)</td>
<td>0.11 [-0.34, 0.56]</td>
</tr>
<tr>
<td>Moynihan et al. (2010)</td>
<td>0.48 [0.07, 0.89]</td>
</tr>
<tr>
<td>Random</td>
<td>0.25 [0.04, 0.46]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0.36 [0.14, 0.59]</strong></td>
</tr>
</tbody>
</table>

Sensitivity analysis in which Banyard et al.’s (2007) one-session intervention was replaced by the three-session intervention did not substantively change these findings. As illustrated in Figure 5, the treatment effect for the five randomized studies (which included Banyard et al., 2007) was significant and positive ($g = 0.28$, 95% CI [0.07, 0.49]) with non-significant heterogeneity ($X^2 = 1.49$ [$p = .47$], $I^2 = 0.00\%$, $\tau^2 = 0$). Across all studies, the average intervention effect (for randomized and non-randomized studies) was significant and positive ($g = 0.38$, 95% CI [0.17, 0.58]) with non-significant heterogeneity ($X^2 = 5.18$ [$p = .39$], $I^2 = 12.40\%$, $\tau^2 = 0.01$).
Figure 5: Forest Plot of Bystander Intervention Effects on Rape Myth Acceptance – One to Four Months Post-Intervention - with Alternative Tx Arm for Banyard et al. (2007)

Table: Rape Myth Acceptance: 1 to 4 Months Post-test

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonrandom Assignment</td>
<td></td>
</tr>
<tr>
<td>Baker et al. (2014)</td>
<td>0.82 [-0.20, 1.84]</td>
</tr>
<tr>
<td>Foubert &amp; Marriott (1997)</td>
<td>0.28 [-0.50, 1.06]</td>
</tr>
<tr>
<td>Peterson et al. (2016)</td>
<td>0.71 [0.26, 1.16]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>0.63 [0.27, 1.00]</td>
</tr>
<tr>
<td>Random Assignment</td>
<td></td>
</tr>
<tr>
<td>Banyard et al. (2007) Alt Tx Arm</td>
<td>0.25 [-0.04, 0.54]</td>
</tr>
<tr>
<td>Gidycz et al. (2011)</td>
<td>0.11 [-0.34, 0.56]</td>
</tr>
<tr>
<td>Moynihan et al. (2010)</td>
<td>0.48 [0.07, 0.89]</td>
</tr>
<tr>
<td>Random</td>
<td>0.28 [0.07, 0.49]</td>
</tr>
<tr>
<td>Total</td>
<td>0.38 [0.17, 0.58]</td>
</tr>
</tbody>
</table>

Six to Seven Month Post-Intervention Effects. Four studies reported rape myth acceptance outcomes at six to seven months post-intervention. Each of these studies involved random assignment of participants to conditions. As shown in the forest plot in Figure 6, the average intervention effect across these four studies was significant and positive ($g = 0.38, 95\%\ CI [0.17, 0.58]$) with non-significant between-study heterogeneity ($X^2 = 5.18 [p = .39], I^2 = 12.40\%, \tau^2 = 0.01$). These findings indicate that bystander programs have a positive (beneficial) effect on rape myth acceptance six to seven months post-intervention. However, the small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
Figure 6: Forest Plot of Bystander Intervention Effects on Rape Myth Acceptance – Six to Seven Months Post-Intervention

<table>
<thead>
<tr>
<th>Study</th>
<th>Rape Myth Acceptance: 6 to 7 Months Post-test</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foubert et al. (2007)</td>
<td></td>
<td>0.01 [-0.17 , 0.19]</td>
</tr>
<tr>
<td>Foubert (2000)</td>
<td></td>
<td>0.23 [-0.40 , 0.86]</td>
</tr>
<tr>
<td>Gidycz et al. (2011)</td>
<td></td>
<td>0.06 [-0.39 , 0.51]</td>
</tr>
<tr>
<td>Salazar et al. (2014)</td>
<td></td>
<td>0.99 [ 0.70 , 1.28]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>0.38 [ 0.17 , 0.58]</strong></td>
</tr>
</tbody>
</table>

**Gender attitudes**

Six studies measured gender attitudes as an intervention outcome. These studies operationalized gender attitudes or gender-role ideology through multiple-item scales consisting of statements such as “If men pay for a date, they deserve something in return.” All but two of these studies (Miller et al., 2014; Peterson et al., 2016) involved random assignment of participants to conditions. All but one study (Brokenshire, 2015) was published in a peer-reviewed outlet; this study was an unpublished master’s thesis. We collapsed gender attitude effect sizes into three timing intervals: (1) immediate post-test (i.e., zero weeks to one week), (2) two to four months post-intervention, and (3) six months to one year post-intervention.

**Immediate Post-test Effects.** One study (Brokenshire, 2015) reported gender attitudes as an immediate post-test. This study indicated that bystander programs had a non-significant but negative effect on gender attitudes at immediate post-test ($g = -0.41$, 95% CI [-0.88, 0.06]).

**Two to Four Month Intervention Effects.** Three studies reported gender attitudes two to four months post-intervention. As shown in the forest plot in Figure 7, the intervention effect for the one non-randomized study was significant and positive ($g = 0.79$, 95% CI [0.32, 1.26]). The average effect for the two randomized studies was non-significant ($g = -0.10$, 95% CI [-0.35, 0.75]) with non-significant heterogeneity ($X^2 = 0.43$ [p = .51], I$^2 = 0.00\%$, $\tau^2 = 0$). Across all studies (randomized and non-randomized), the average intervention effect was non-significant ($g = 0.20$, 95% CI [-0.35, 0.75]) with significant between-study heterogeneity ($X^2 = 12.02$ [p = .00], I$^2 = 84.17\%$, $\tau^2 = 0.20$). Thus, these findings indicate that there is no evidence that bystander programs have an effect on gender attitudes two to four months post-intervention. Again, however, the small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
### Six Month to One-Year Intervention Effects

Four studies reported gender attitudes six months to one year post-intervention. As shown in the forest plot in Figure 8, the intervention effect for the one non-randomized study was non-significant ($g = 0.05$, 95% CI $[-0.30, 0.40]$). The intervention effect for the three randomized studies was non-significant ($g = 0.19$, 95% CI $[-0.32, 0.70]$) with significant heterogeneity ($X^2 = 17.15$ [\(p = .00\)], $I^2 = 85.94\%$, $\tau^2 = 0.17$). Across all studies (randomized and non-randomized), the average intervention effect was non-significant ($g = 0.16$, 95% CI $[-0.22, 0.53]$) with significant heterogeneity ($X^2 = 18.00$ [\(p = .00\)], $I^2 = 80.07\%$, $\tau^2 = 0.12$). Thus, there is no evidence that bystander programs had an effect on gender attitude outcomes six months to one year post-intervention. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
Date rape attitudes

Four studies measured data rape attitudes. These studies operationalized date rape attitudes with multi-item scales such as the College Date Rate Attitude Survey (Lanier & Elliot, 1997) and the Rape Attitudes and Beliefs Scale (Burgess, 2007). Such scales asked respondents to rate their approval of statements such as, “Many women pretend they don’t want to have sex because they don’t want to appear ‘easy.’” All four of these studies involved random assignment of participants to conditions. Two of these studies were published in peer-reviewed outlets (Banyard et al., 2007; Salazar et al., 2014) and two were unpublished theses/dissertations (Brokenshire, 2015; Chiriboga, 2016). More than one study reported immediate post-test effects, so we synthesized these together. The timing of follow-up waves was too disparate between studies to permit synthesis; thus, we report these results individually.

**Immediate Post-Test Effects.** Three studies reported date rape attitudes as an immediate post-test outcome. As shown in Figure 9, there was no evidence that bystander interventions had an effect on date rape attitudes at immediate follow-up ($g = 0.04$, 95% CI [-0.29, 0.38]; $X^2 = 5.68 \ [p = .06]$; $I^2 = 64.81\%$, $r^2 = .06$). The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
Sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention with the three-session intervention did not substantively change these findings ($g = 0.07$, 95% CI [-0.30, 0.45]) and indicated significant between-study heterogeneity in effects ($X^2 = 7.07$ [p = .03], $I^2 = 70.93\%$, $\tau^2 = .08$). See Figure 10.
Follow-Up Effects. Two studies reported a date rape attitude outcome beyond the immediate post-test. The timing of these measures was too disparate to permit synthesis of effect sizes. Banyard et al. (2007) reported date rape attitudes two months post-intervention for both the (focal) one-session intervention ($g = 0.23$, 95% CI $[-0.06, 0.52]$) and the (alternative) three-session intervention ($g = 0.21$, 95% CI $[-0.06, 0.48]$). Neither of these intervention effects was statistically significant. Salazar et al. (2014) reported date rape attitudes six months post-intervention and reported a positive and statistically significant intervention effect ($g = 0.91$, 95% CI $[0.64, 1.18]$).

Noticing a Sexual Assault or its Warning Signs

Four studies reported a measure of whether participants noticed a sexual assault occurring. Studies operationalized this outcome in a number of ways, including a single item developed by Burn (2009): “At a party or bar, I am probably too busy to be aware of whether someone is at risk for sexually assaulting someone.” All but one of these studies (Senn & Forrest, 2016) involved random assignment of participants to conditions and all but one (Brokenshire, 2015) were published in a peer-reviewed outlet. This specific report was an unpublished master’s thesis. We collapsed noticing sexual assault effect sizes into three timing intervals: (1) immediate post-test (i.e., zero weeks to one week), (2) one month to four months, and (3) one year.

Immediate Post-Test. Two studies reported noticing sexual assault as an immediate post-test outcome (i.e., zero weeks to one week post intervention). As shown in the forest plot in Figure 11, the intervention effect for the one non-randomized study was significant and positive ($g = 0.26$, 95% CI $[0.06, 0.46]$). Conversely, the effect for the one randomized study was non-significant ($g = -0.14$, 95% CI $[-0.81, 0.53]$). Across the two studies, the average intervention effect was non-significant ($g = 0.19$, 95% CI $[-0.10, 0.45]$) with non-significant heterogeneity ($X^2 = 1.27$ [$p = .26$],
Thus, at immediate post-intervention, there was no evidence that bystander programs have an effect on participants’ reports of noticing a sexual assault. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.

**Figure 11: Forest Plot of Bystander Intervention Effects on Noticing Sexual Assault – Immediate Post-Test**

<table>
<thead>
<tr>
<th>Study</th>
<th>Noticing Sexual Assault: Immediate Post-test</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonrandom Assignment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td></td>
<td>0.26 [0.06 , 0.46]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td></td>
<td>0.26 [0.06 , 0.46]</td>
</tr>
<tr>
<td>Random Assignment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brokenshire (2015)</td>
<td></td>
<td>-0.14 [-0.81 , 0.53]</td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td>-0.14 [-0.81 , 0.53]</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0.19 [-0.10 , 0.49]</td>
</tr>
</tbody>
</table>

**One to Four-Month Follow-Up.** Three studies reported noticing sexual assault as an outcome one to four-months post-intervention. As shown in the forest plot in Figure 12, the intervention effect for the one non-randomized study was non-significant ($g = 0.11, 95\% \text{ CI} [-0.09, 0.31]$). The average effect for the two randomized studies was also non-significant ($g = -0.04, 95\% \text{ CI} [-0.26, 0.17]$) with non-significant heterogeneity ($X^2 = 0.60 \ [p = .44], I^2 = 0.00\% , \tau^2 = 0$). Across all studies (randomized and non-randomized), the average intervention effect was non-significant ($g = 0.04, 95\% \text{ CI} [-0.10, 0.19]$) with non-significant heterogeneity ($X^2 = 1.66 \ [p = .44], I^2 = 0.00\% , \tau^2 = 0$). Thus, there was no evidence that bystander programs have an effect on participants’ reports of noticing a sexual assault one to four-months post-intervention. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
Figure 12: Forest Plot of Bystander Intervention Effects on Noticing Sexual Assault – One to Four Months Post-Intervention

<table>
<thead>
<tr>
<th>Study</th>
<th>Noticing Sexual Assault: 1 to 4 Months Post-test</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td></td>
<td>0.11 [-0.09, 0.31]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td></td>
<td>0.11 [-0.09, 0.31]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller et al. (2013)</td>
<td></td>
<td>-0.01 [-0.25, 0.23]</td>
</tr>
<tr>
<td>Moynihan et al. (2011)</td>
<td></td>
<td>-0.26 [-0.85, 0.33]</td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td>-0.04 [-0.26, 0.17]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>0.04 [-0.10, 0.19]</td>
</tr>
</tbody>
</table>

One Year Follow-Up. One study reported noticing sexual assault as an intervention outcome one year post-intervention. The effect size from this study (Miller at al., 2013) was non-significant ($g = -0.06$, 95% CI [-0.13, 0.41]), thus providing no evidence that the program in this study affected respondents’ noticing a sexual assault one year post-intervention.

Identifying a Situation as Appropriate for Intervention

Six studies reported a measure of participants’ identification of a situation as appropriate for intervention. Studies frequently operationalized this concept using some adaptation of Burn’s (2009) Failure to Identify Situation as High Risk scale (e.g., “In a party or bar situation, I think I might be uncertain as to whether someone is at-risk for being sexually assaulted”). All but three of these studies (Amar et al., 2015; Baker et al., 2014; Senn & Forrest, 2016) involved random assignment of participants to conditions and all but one (Addison, 2015) was published in a peer-reviewed outlet. This specific report was an unpublished doctoral dissertation. We collapsed noticing sexual assault effect sizes into three timing intervals: (1) immediate post-test (i.e., zero weeks to one week), (2) one month to four months post-intervention, and (3) six months post-intervention.

Immediate Post-Test. Five studies reported identification of a situation as appropriate for intervention as an immediate post-test outcome (i.e., zero weeks to one week post intervention). As shown in the forest plot in Figure 13, the average intervention effect for the non-randomized studies was non-significant ($g = 0.43$, 95% CI [-0.43, 1.30]) with significant heterogeneity across studies ($X^2 = 15.74 \ [p = .00], I^2 = 91.24\%, \tau^2 = 0.49$). The average intervention effect for the randomized studies was significant and positive ($g = 0.76$, 95% CI [0.48, 1.05]), however, with non-significant heterogeneity ($X^2 = 1.58 \ [p = .21], I^2 = 36.78\%, \tau^2 = 0.02$). Across all studies (randomized and non-randomized), the average intervention effect was significant and positive ($g = 0.57$, 95% CI [0.08, 1.05]) with significant heterogeneity across studies ($X^2 = 23.31 \ [p = .00], I^2 = 89.71\%, \tau^2 = 0.25$). Thus, at immediate post-intervention, bystander programs have a positive
(beneficial) and significant beneficial effect on participants’ identification of a situation as appropriate for intervention. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.

**Figure 13: Forest Plot of Bystander Intervention Effects on Identifying a Situation as Appropriate for Intervention – Immediate Post-Test**

<table>
<thead>
<tr>
<th>Study</th>
<th>Identifying Situation: Immediate Post-test</th>
<th>Hedges $g$ [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amar et al. (2015)</td>
<td></td>
<td>-0.30 [-0.71, 0.11]</td>
</tr>
<tr>
<td>Baker et al. (2014)</td>
<td></td>
<td>1.40 [ 0.30, 2.50 ]</td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td></td>
<td>0.52 [ 0.32, 0.72 ]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td></td>
<td>0.43 [-0.43, 1.30 ]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addison (2015)</td>
<td></td>
<td>0.95 [ 0.56, 1.34 ]</td>
</tr>
<tr>
<td>Banyard et al. (2007)</td>
<td></td>
<td>0.65 [ 0.40, 0.90 ]</td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td>0.76 [ 0.48, 1.05 ]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>0.57 [ 0.08, 1.05 ]</td>
</tr>
</tbody>
</table>

We conducted a sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention arm with the three-session intervention arm. As shown in Figure 14, findings were substantively similar to those from the main analysis. The intervention effect for the two randomized studies (which included Banyard et al.) was still significant and positive ($g = 1.23$, 95% CI [0.71, 1.75]; $X^2 = 4.49$ [$p = .03$], $I^2 = 77.75\%$, $\tau^2 = 0.011$). The average intervention effect across all studies was also still significant and positive ($g = 0.77$, 95% CI [0.12, 1.42]; $X^2 = 56.05$ [$p = .00$], $I^2 = 93.82\%$, $\tau^2 = 0.47$). Importantly, the replacement of Banyard et al.’s one-session intervention arm with their three-session intervention arm increased the magnitude of the average intervention effect for the randomized studies (0.76 to 1.23) as well as for the total sample (0.57 to 0.77).
**Identifying Situation: Immediate Post-test**

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Amar et al. (2015)</td>
<td>-0.30 [-0.71, 0.11]</td>
</tr>
<tr>
<td>Baker et al. (2014)</td>
<td>1.40 [0.30, 2.50]</td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td>0.52 [0.32, 0.72]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>0.43 [-0.43, 1.30]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Addison (2015)</td>
<td>0.95 [0.56, 1.34]</td>
</tr>
<tr>
<td>Banyard et al. (2007) - Alt Tx Arm</td>
<td>1.48 [1.19, 1.77]</td>
</tr>
<tr>
<td>Random</td>
<td>1.23 [0.71, 1.75]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.77 [0.12, 1.42]</td>
</tr>
</tbody>
</table>

**One to Four-Month Follow-Up.** Three studies reported identification of a situation as appropriate for intervention one to four-months post-intervention. The effect size for one of these studies (Senn & Forrest, 2016) fell more than two standard deviations below the mean of the distribution for the total sample ($g = 0.26$) so we Winsorized it by replacing it with the value that fell exactly two standard deviations below the mean ($g = 0.33$). As shown in the forest plot in Figure 15, the average intervention effect for the two non-randomized studies was non-significant ($g = 0.58, 95\%$ CI [-0.17, 1.33]) with non-significant heterogeneity ($X^2 = 2.26 [p = .13], I^2 = 55.71\%, \tau^2 = 0.20$). The intervention effect for the one randomized study was significant and positive ($g = 0.46, 95\%$ CI [0.17, 0.75]). Across all studies (randomized and non-randomized), the average intervention effect was significant and positive ($g = 0.39, 95\%$ CI [0.23, 0.55]) with non-significant heterogeneity across studies ($X^2 = 2.59 [p = .27], I^2 = 0.01\%, \tau^2 = 0$). Thus, the overall mean effect size indicates that bystander programs have a significant beneficial effect on participants’ identification of a situation as appropriate for intervention one-to-four months post-intervention. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
We conducted a sensitivity analysis to determine whether Winsorizing the effect size for Senn and Forrest (2016) affected our overall findings. Running the meta-analysis with the original effect size for this study produced nearly identical findings to those from the main analysis for both the two non-randomized studies ($g = 0.55$, 95% CI [-0.28, 1.39]) and the total sample containing both randomized and non-randomized studies ($g = 0.37$, 95% CI [0.15, 0.59]). Tests for heterogeneity revealed non-significant heterogeneity, although specific values were slightly different than those from the main analysis for both the non-randomized studies ($X^2 = 2.65 \ [p = .10], I^2 = 62.26\%, \tau^2 = 0.26$) and the total sample ($X^2 = 3.56 \ [p = .17], I^2 = 26.48\%, \tau^2 = 0.01$).

We also conducted a sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention arm with the three-session intervention arm. Inspection of the distribution of effect sizes revealed that there were no outliers in the total sample containing Banyard et al.’s three-session intervention arm. As indicated in Figure 16, findings were substantively similar to those from the main analysis. The intervention effect for Banyard et al. (2007), the one randomized study in the sample, remained significant and positive ($g = 1.07$, 95% CI [0.78, 1.36]). The overall average effect for randomized and non-randomized studies) also remained significant and positive ($g = 0.75$, 95% CI [0.13, 1.37]) with significant heterogeneity across studies ($X^2 = 21.60 \ [p = .00], I^2 = 88.54\%, \tau^2 = 0.23$).
Six Month Follow-Up. One study reported participants’ identification of a situation as appropriate for intervention as an outcome six months post-intervention. The effect size from this study (Salazar et al., 2014) was significant and positive ($g = 1.56$, 95% CI [1.25, 1.87]).

Taking Responsibility for Acting/Intervening

Four studies reported a measure of participants’ taking responsibility for acting/intervening when witnessing violence or its warning signs. Studies operationalized this outcome using either the full or an adapted version of the Failure to Take Intervention Responsibility scale developed by Burn (2009), which consisted of items such as “I am less likely to intervene to reduce a person’s risk of sexual assault if I think she/he made choices that increased their risk.” Two studies (Banyard et al., 2007; Moynihan et al., 2011) involved random assignment of participants to conditions and two (Amar et al., 2015, Senn & Forest, 2016) did not. All four studies were published in a peer-reviewed outlet. We collapsed noticing sexual assault effect sizes into two timing intervals: (1) immediate post-test (i.e., zero weeks to one week) and (2) one month to four months post-intervention.

Immediate Post-Test. Three studies reported taking responsibility for acting/intervening as an immediate post-test outcome (i.e., zero weeks to one week post intervention). As shown in the forest plot in Figure 17, the intervention effect for the two non-randomized studies was non-significant ($g = -0.05$, 95% CI [-0.90, 0.81]) with significant heterogeneity across studies ($X^2 = 14.00 \ [p = .00], I^2 = 92.85\%$, $\tau^2 = 0.35$). The intervention effect for the randomized study was significant and positive ($g = 0.96$, 95% CI [0.69, 1.23]). The overall average intervention effect (for the randomized and non-randomized studies) was non-significant ($g = 0.29$, 95% CI [-0.53, 1.11]) with significant heterogeneity ($X^2 = 34.30 \ [p = .00], I^2 = 96.07\%$, $\tau^2 = 0.50$). The overall average intervention effect indicates that, at immediate post-intervention, there is no evidence that bystander programs have an effect on participants’ taking responsibility for acting/intervening.
when witnessing violence or its warning signs. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.

**Figure 17: Forest Plot of Bystander Intervention Effects on Taking Responsibility to Act/Intervene – Immediate Post-Test**

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges $g$ [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Amar et al. (2015)</td>
<td>-0.50 [-0.91, -0.09]</td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td>0.37 [0.17, 0.57]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>-0.05 [-0.90, 0.81]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Banyard et al. (2007)</td>
<td>0.96 [0.69, 1.23]</td>
</tr>
<tr>
<td>Random</td>
<td>0.96 [0.69, 1.23]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.29 [-0.53, 1.11]</td>
</tr>
</tbody>
</table>

We conducted a sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention arm with the three-session intervention arm. As indicated in Figure 18, findings were substantively similar to those from the main analysis. The intervention effect for Banyard et al. (2007), the single randomized study, was significant and positive ($g = 1.51, 95\% \text{ CI } [1.22, 1.80]$). The overall average intervention effect (for randomized and non-randomized studies) was non-significant ($g = 0.47, 95\% \text{ CI } [-0.67, 1.60]$) with significant heterogeneity across studies ($X^2 = 69.20 \ [p = .00], I^2 = 97.83\%, \tau^2 = 0.98$).
One to Four-Month Follow-Up. Three studies reported taking responsibility for acting/intervening one to four-months post-intervention. As shown in the forest plot in Figure 19, the intervention effect for the one non-randomized study was non-significant ($g = 0.16$, 95% CI [-0.04, 0.36]). The average intervention effect for the two randomized studies was significant and positive ($g = 0.50$, 95% CI [0.24, 0.76]) with non-significant heterogeneity ($X^2 = 0.60$ [$p = .44$], $I^2 = 0.00\%$, $\tau^2 = 0$). Overall, the average intervention effect (for the randomized and non-randomized studies) was significant and positive ($g = 0.32$, 95% CI [0.04, 0.61]) with non-significant heterogeneity ($X^2 = 4.68$ [$p = .10$], $I^2 = 56.86\%$, $\tau^2 = 0.03$). Thus, the results indicate that bystander programs have a significant positive (beneficial) effect on participants’ taking responsibility for acting/intervening when witnessing violence or its warning signs one to four-months post-intervention. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
We conducted a sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention arm with the three-session intervention arm. As indicated in Figure 20, substituting this intervention arm resulted in substantively different findings from those in the main analysis. The intervention effect for the two randomized studies (which included Banyard et al., 2007) remained significant and positive ($g = 0.67$, 95% CI [0.04, 1.30]) with non-significant heterogeneity ($X^2 = 3.76$ [$p = .05$], $I^2 = 73.37\%$, $\tau^2 = 0.16$). However, the overall average intervention effect (for randomized and non-randomized studies) was non-significant ($g = 0.47$, 95% CI [-0.04, 0.98]) with significant heterogeneity across studies ($X^2 = 18.84$ [$p = .00$], $I^2 = 86.59\%$, $\tau^2 = 0.17$).
Knowing Strategies for Helping/Intervening

Four studies reported a measure of participants’ knowledge of strategies for helping/intervening. Studies operationalized this concept using Burn’s (2009) two-item Failure to Intervene Due to Skills Deficit scale, which asked respondents to rate their agreement with statements such as “Although I would like to intervene when a guy’s sexual conduct is questionable, I am not sure I would know what to say or do.” Two studies (Brokenshire, 2015; Potter et al., 2008) involved random assignment of participants to conditions and two (Amar et al., 2015, Senn & Forest, 2016) did not. All but one study (Brokenshire, 2015) was published in a peer-reviewed outlet. This specific report was an unpublished master’s thesis. We collapsed knowing strategies for helping/intervening into two timing intervals: (1) immediate post-test (i.e., zero weeks to two weeks) and (2) four months post-intervention.

Immediate Post-Test. Four studies reported knowing strategies for helping/intervening as an immediate post-test outcome (i.e., zero weeks to two weeks post intervention). As shown in the forest plot in Figure 21, the average intervention effect for the two non-randomized studies was non-significant \( g = 0.33, 95\% \text{ CI} [-0.70, 1.36] \) with significant heterogeneity across studies \( (X^2 = 20.38 [p = .00], I^2 = 95.10\% , \tau^2 = 0.52) \). The treatment effect for the randomized studies was non-significant \( g = 0.83, 95\% \text{ CI} [-0.03, 1.69] \) with non-significant heterogeneity \( (X^2 = 2.29 [p = .13], I^2 = 56.32\% , \tau^2 = 0.22) \). Overall, the average intervention effect (for the randomized and non-randomized studies) was non-significant \( g = 0.54, 95\% \text{ CI} [-0.09, 1.17] \) with significant heterogeneity \( (X^2 = 23.50 [p = .00], I^2 = 86.46\% , \tau^2 = 0.33) \). Thus, at immediate post-intervention, there is no evidence that bystander programs have an effect on participants’ knowledge of strategies for helping/intervening. The small sample size in this meta-analysis \( (n < 10) \) precluded ad-hoc analysis of moderators or small study/publication bias.
Four-Month Follow-Up. One study (Senn & Forrest, 2016) reported knowing strategies for helping/intervening four months post-intervention. The effect size for this study was significant and positive ($g = 0.63$, 95% CI [0.43, 0.83]).

Bystander efficacy

Eleven studies reported a measure of bystander efficacy as an outcome. These studies operationalized bystander efficacy using a modified or adapted version of the Bystander Efficacy Scale (Banyard et al., 2007). This scale asked participants to rate their confidence in their ability to perform behaviors such as “try and stop or discourage someone who is spreading rumors online about another person’s body or sexual behavior” or “do something to help a very drunk person who is being brought upstairs to a bedroom by a group of people at a party.” All but three of these studies (Baker et al., 2014; Peterson et al., 2016; Senn & Forrest, 2016) involved random assignment of participants to conditions and all but one (Addison, 2015) was published in a peer-reviewed outlet. This specific report was an unpublished dissertation. We collapsed bystander efficacy effect sizes into three timing intervals: (1) immediate post-test (i.e., zero weeks to one week), (2) one month to four months post-intervention, and (3) six months post-intervention.

Immediate Post-Test. Eight studies reported bystander efficacy as an immediate post-test outcome (i.e., zero weeks to one week post intervention). As shown in the forest plot in Figure 22, the average intervention effect for the two non-randomized studies was significant and positive ($g = 0.23$, 95% CI [0.04, 0.43]) with non-significant heterogeneity ($X^2 = 0.53$ [$p = .46$], $I^2 = 0.00\%$, $\tau^2 = 0$). The average intervention effect for the six randomized studies was significant and positive ($g = 0.49$, 95% CI [0.27, 0.72]) with significant heterogeneity across individual studies ($X^2 = 18.48$ [$p = .00$], $I^2 = 70.51\%$, $\tau^2 = 0.05$). Across all studies (randomized and non-randomized), the average
The intervention effect was significant and positive ($g = 0.45$, 95% CI [0.25, 0.65]) with significant heterogeneity across studies ($X^2 = 25.17 \ [p = .00], I^2 = 69.78\%$, $\tau^2 = 0.05$). Thus, at immediate post-intervention, bystander programs had a significant and positive effect on bystander efficacy. The small sample size in this meta-analysis ($n<10$) precluded ad-hoc analysis of moderators or small study/publication bias.

**Figure 22: Forest Plot of Bystander Intervention Effects on Bystander Efficacy – Immediate Post-Test**

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Baker et al. (2014)</td>
<td>0.60 [-0.40, 1.60]</td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td>0.22 [0.02, 0.42]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>0.23 [0.04, 0.43]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Addison (2015)</td>
<td>0.89 [0.65, 1.13]</td>
</tr>
<tr>
<td>Banyard et al. (2007)</td>
<td>0.51 [0.26, 0.76]</td>
</tr>
<tr>
<td>Jouriles et al. (2016a)</td>
<td>0.41 [0.14, 0.68]</td>
</tr>
<tr>
<td>Jouriles et al. (2016b)</td>
<td>0.10 [-0.19, 0.39]</td>
</tr>
<tr>
<td>Kleinssasser et al. (2015)</td>
<td>0.40 [-0.01, 0.81]</td>
</tr>
<tr>
<td>Moynihan et al. (2010)</td>
<td>0.61 [0.20, 1.02]</td>
</tr>
<tr>
<td>Random</td>
<td>0.49 [0.27, 0.72]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.45 [0.25, 0.65]</td>
</tr>
</tbody>
</table>

We conducted a sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention arm with the three-session intervention arm. As indicated in Figure 23, findings were substantively similar to those from the main analysis. The average intervention effect for the six randomized studies (which included Banyard et al.) was significant and positive ($g = 0.54$, 95% CI [0.30, 0.78]) with significant heterogeneity ($X^2 = 21.05 \ [p = .00], I^2 = 74.41\%, \tau^2 = 0.07$). The overall average intervention effect (for randomized and non-randomized studies) was also significant and positive ($g = 0.49$, 95% CI [0.27, 0.71]) with significant heterogeneity across studies ($X^2 = 30.32 \ [p = .00], I^2 = 74.30\%, \tau^2 = 0.07$).
**Figure 23: Forest Plot of Bystander Intervention Effects on Bystander Efficacy – Immediate Post-Test with Alternative Tx Arm for Banyard et al. (2007)**

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Baker et al. (2014)</td>
<td>0.60 [-0.40, 1.60]</td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td>0.22 [0.02, 0.42]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>0.23 [0.04, 0.43]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Addison (2015)</td>
<td>0.89 [0.65, 1.13]</td>
</tr>
<tr>
<td>Banyard et al. (2007) Alt Tx Arm</td>
<td>0.76 [0.51, 1.01]</td>
</tr>
<tr>
<td>Jouriles et al. (2016a)</td>
<td>0.41 [0.14, 0.68]</td>
</tr>
<tr>
<td>Jouriles et al. (2016b)</td>
<td>0.10 [-0.19, 0.39]</td>
</tr>
<tr>
<td>Kleinsasser et al. (2015)</td>
<td>0.40 [-0.01, 0.81]</td>
</tr>
<tr>
<td>Moynihan et al. (2010)</td>
<td>0.61 [0.20, 1.02]</td>
</tr>
<tr>
<td>Random</td>
<td>0.54 [0.30, 0.78]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0.49 [0.27, 0.71]</strong></td>
</tr>
</tbody>
</table>

**One to Four-Month Follow-Up.** Nine studies reported bystander efficacy one to four months post-intervention. As shown in Figure 24, the average intervention effect for the three non-randomized studies was non-significant (\(g = 0.50, 95\% \text{ CI } [-0.07, 1.06]\)) with significant heterogeneity (\(X^2 = 10.03 [p = .01], I^2 = 76.18\%, \tau^2 = 0.18\)). The average intervention effect for the six randomized studies was significant and positive (\(g = 0.52, 95\% \text{ CI } [0.36, 0.68]\)) with non-significant heterogeneity (\(X^2 = 5.65 [p = .34], I^2 = 20.50\%, \tau^2 = 0.01\)). Across all studies (randomized and non-randomized), the average intervention effect was significant and positive (\(g = 0.50, 95\% \text{ CI } [0.31, 0.68]\)) but with significant heterogeneity (\(X^2 = 20.19 [p = .01], I^2 = 58.56\%, \tau^2 = 0.04\)). Thus, bystander programs have a significant positive (beneficial) effect on bystander efficacy at one to four-months post-intervention. The small sample size in this meta-analysis (\(n < 10\)) precluded ad-hoc analysis of moderators or small study/publication bias.
We conducted a sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention arm with the three-session intervention arm. As indicated in Figure 25, findings were substantively similar to those from the main analysis. The average intervention effect for the six randomized studies (which included Banyard et al.) was significant and positive ($g = 0.53$, 95% CI [0.37, 0.69]) with non-significant heterogeneity ($X^2 = 5.68$ [p = .34], $I^2 = 20.90\%$, $\tau^2 = 0.01$).

Across all studies (randomized and non-randomized), the average intervention effect was also significant and positive ($g = 0.50$, 95% CI [0.31, 0.69]) but with significant heterogeneity across studies ($X^2 = 20.66$ [p = .01], $I^2 = 59.10\%$, $\tau^2 = 0.04$).
**Six Month Follow-Up.** One study (Salazar et al., 2014) reported bystander efficacy as an outcome six months post-intervention. Since this study measured bystander efficacy two to six months later than all other studies reporting this outcome, we elected to report it separately. Findings from this particular study indicated a non-significant program effect ($g = 0.13, 95\% \text{ CI } [-0.14, 0.40]$) six months post-intervention.

**Bystander intentions**

Eleven studies reported a measure of bystander intentions as an outcome. Studies operationalized this outcome using the full or adapted Bystander Intent to Help Scale (Banyard, Moynihan, Cares, & Warner, 2014), which asks participants their likelihood to engage in bystander behavior (e.g., “I stop and check in on someone who looks intoxicated when they are being taken upstairs at a party.”). All but four of these studies (Amar et al., 2014; Miller et al., 2014; Peterson et al., 2016; Senn & Forrest, 2016) involved random assignment of participants to conditions and all but one (Brokenshire, 2015) was published in a peer-reviewed outlet. This specific report was an unpublished master’s thesis. We collapsed bystander efficacy effect sizes into three timing intervals: (1) immediate post-test (i.e., zero weeks to one week), (2) one month to four months post-intervention, and (3) six months to one year post-intervention.

**Immediate Post-Test.** Six studies reported bystander intentions as an immediate post-test outcome (i.e., zero weeks to one week post intervention). As shown in the forest plot in Figure 26, the average intervention effect for the two non-randomized studies was non-significant ($g = -0.15, 95\% \text{ CI } [-1.04, 0.74]$) but with significant heterogeneity across studies ($X^2 = 15.31 \ [p = .00], F^2 = 93.47\%, \tau^2 = 0.39$). The average intervention effect for the four randomized studies was significant
and positive ($g = 0.32, 95\% \text{ CI} [0.01, 0.64]$) with significant heterogeneity ($X^2 = 10.94 \ [p = .01], I^2 = 73.17\%, \tau^2 = 0.07$). Across all studies (randomized and non-randomized), the average intervention effect was non-significant ($g = 0.17, 95\% \text{ CI} [-0.18, 0.52]$) but with significant heterogeneity ($X^2 = 30.25 \ [p = .00], I^2 = 87.57\%, \tau^2 = 0.16$). The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.

Interestingly, the average intervention effect for the randomized studies was significant and positive, but when synthesized with the two non-randomized studies, the average intervention effect was non-significant. This attenuation appears to be driven by the extreme (negative) effect size reported by Amar et al. (2015). However, preliminary inspection of the distribution of effect sizes indicated this was not an outlier; thus, we did not Winsorize it in our main analysis.

Nevertheless, in light of these finding we ran a sensitivity analysis in which we dropped Amar et al. (2015) from the sample. When removing Amar et al. (2015), the overall average effect size was significant and positive ($g = 0.32, 95\% \text{ CI} [0.08, 0.56]$) with significant heterogeneity ($X^2 = 11.26 \ [p = .02], I^2 = 67.19\%, \tau^2 = 0.04$).

---

Figure 26: Forest Plot of Bystander Intervention Effects on Bystander Intentions – Immediate Post-Test

We conducted a sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention arm with the three-session intervention arm. As indicated in Figure 27, findings were substantively similar to those from the main analysis. The average intervention effect for the six randomized studies (which included Banyard et al.) was non-significant ($g = 0.36, 95\% \text{ CI} [-0.01, 0.73]$) with significant heterogeneity across studies ($X^2 = 16.34 \ [p = .00], I^2 = 81.16\%, \tau^2 = 0.12$). Across all studies (randomized and non-randomized), the average intervention effect was non-significant ($g = 0.19, 95\% \text{ CI} [-0.19, 0.57]$) but with significant heterogeneity across individual study effect sizes ($X^2 = 37.74 \ [p = .00], I^2 = 89.59\%, \tau^2 = 0.20$).
One to Four-Month Follow-Up. Six studies reported a measure of bystander intentions one to four months post-intervention. As shown in Figure 28, the average intervention effect for the non-randomized studies was non-significant ($g = 0.51$, 95% CI [-0.40, 1.42]) with significant heterogeneity across individual effect sizes ($X^2 = 13.16$ [$p = .00$], $I^2 = 92.40\%$, $\tau^2 = 0.40$). The average intervention effect for the randomized studies was significant and positive ($g = 0.44$, 95% CI [0.24, 0.63]) with non-significant heterogeneity ($X^2 = 1.77$ [$p = .62$], $I^2 = 0.00\%$, $\tau^2 = 0$). Across all studies (randomized and non-randomized), the average intervention effect was significant and positive ($g = 0.41$, 95% CI [0.15, 0.68]) but with significant heterogeneity across studies ($X^2 = 18.64$ [$p = .00$], $I^2 = 70.16\%$, $\tau^2 = 0.07$). Overall, these findings indicate bystander programs have a significant positive (beneficial) effect on bystander intentions one- to four-months post-intervention. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
We conducted a sensitivity analysis in which we replaced Banyard et al.’s (2007) one-session intervention arm with the three-session treatment arm. As indicated in Figure 29, findings were substantively similar to those from the main analysis. The average intervention effect for the four randomized studies (which included Banyard et al.) was significant and positive ($g = 0.41$, 95% CI $[0.21, 0.61]$) with non-significant heterogeneity effect sizes ($X^2 = 1.47 \ [p = .69]$, $I^2 = 0.00\%$, $\tau^2 = 0$). Across all studies, the average intervention effect was significant and positive ($g = 0.40$, 95% CI $[0.14, 0.67]$) with significant heterogeneity ($X^2 = 17.60 \ [p = .00]$, $I^2 = 69.26\%$, $\tau^2 = 0.07$).
**Six Months to One Year.** Three studies reported bystander intentions six months to one year post-intervention. As shown in Figure 30, the program effect for the one non-randomized study was non-significant ($g = 0.11$, 95% CI [-0.24, 0.46]). The average intervention effect for the two randomized studies was significant and positive ($g = 0.29$, 95% CI [0.05, 0.53]) with non-significant heterogeneity across individual effect sizes ($X^2 = 0.03$ [$p = .87$], $I^2 = 0.00\%$, $\tau^2 = 0$). Across all studies (randomized and non-randomized), the average intervention effect was significant and positive ($g = 0.23$, 95% CI [0.03, 0.43]) with non-significant heterogeneity ($X^2 = 0.70$ [$p = .70$], $I^2 = 0.00\%$, $\tau^2 = 0$). Overall, these findings indicate bystander programs had a significant positive (beneficial) effect on bystander intentions six months to one year post-intervention.

---

### Bystander Intentions: 1 to 4 Month Post-test

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Peterson et al. (2016)</td>
<td>1.00 [0.53, 1.47]</td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td>0.07 [-0.11, 0.25]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>0.51 [-0.40, 1.42]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Banyard et al. (2007) Alt Tx Arm</td>
<td>0.45 [0.16, 0.74]</td>
</tr>
<tr>
<td>Miller et al. (2013)</td>
<td>0.28 [-0.15, 0.71]</td>
</tr>
<tr>
<td>Moynihan et al. (2010)</td>
<td>0.56 [0.15, 0.97]</td>
</tr>
<tr>
<td>Moynihan et al. (2011)</td>
<td>0.19 [-0.40, 0.78]</td>
</tr>
<tr>
<td>Random</td>
<td>0.41 [0.21, 0.61]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.40 [0.14, 0.67]</td>
</tr>
</tbody>
</table>

---

*Figure 29: Forest Plot of Bystander Intervention Effects on Bystander Intentions – One to Four Months Post-Intervention with Alternative Tx Arm for Banyard et al. (2007)*

*Figure 30: Forest Plot of Bystander Intervention Effects on Bystander Intentions – Six Months to One Year.*
Bystander intervention

Thirteen studies reported a measure of bystander intervention as an outcome. All studies used a form of the Bystander Behaviors Scale (Banyard et al., 2007), which asked participants to indicate the extent to which they have engaged in bystander behavior (e.g., “walked a friend home from a party who has had too much to drink”). All but three of these studies (Miller et al., 2014; Peterson et al., 2016; Senn & Forrest, 2016) involved random assignment of participants to conditions and all but one (Jouriles et al., n.d.) was published in a peer-reviewed outlet. This specific report was a draft of a manuscript under review for publication at the time of this review.

No studies measured bystander intervention at immediate post-intervention. We thus collapsed bystander intervention effect sizes into two timing intervals: (1) one to four months post-intervention and (2) six months to one year post-intervention.

One to Four-Month Follow-Up. Eleven studies reported bystander behavior one- to four-months post-intervention. Inspection of the distribution of effect sizes revealed that the effect size of one of the non-randomized studies (Peterson et al., 2016) fell more than two standard deviations above the mean of the distribution. We therefore Winsorized this effect size ($g = 0.63$) by replacing it with the value that fell exactly two standard deviations above the mean ($g = 0.60$). As depicted in the forest plot in Figure 31 the average intervention effect for the two non-randomized studies was significant and positive ($g = 0.43$, 95% CI [0.25, 0.61]) with non-significant heterogeneity ($X^2 = 0.59 \ [p = .44], I^2 = 0.00\%, \tau^2 = 0$). The average intervention effect for the nine randomized studies was also significant and positive ($g = 0.23$, 95% CI [0.13, 0.33]) with non-significant heterogeneity ($X^2 = 5.44 \ [p = .71], I^2 = 0.00\%, \tau^2 = 0$). Across all studies (randomized and non-randomized), the average intervention effect was significant and positive ($g = 0.27$, 95% CI [0.19, 0.36]) with non-significant heterogeneity ($X^2 = 9.70 \ [p = .47], I^2 = 2.16\%, \tau^2 = 0$). These findings indicate that, at
one to four-months post-intervention bystander programs have a significant positive (beneficial) effect on bystander intervention.

**Figure 31: Forest Plot of Bystander Intervention Effects on Bystander Intervention – One to Four Months Post-intervention**

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Peterson et al. (2016)</td>
<td>0.60 [ 0.13 , 1.07 ]</td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td>0.40 [ 0.20 , 0.60 ]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>0.43 [ 0.25 , 0.61 ]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Banyard et al. (2007)</td>
<td>0.42 [ 0.13 , 0.71 ]</td>
</tr>
<tr>
<td>Gidycz et al. (2011)</td>
<td>0.16 [ -0.29 , 0.61 ]</td>
</tr>
<tr>
<td>Jouriles et al. (n.d.)</td>
<td>0.26 [ 0.04 , 0.48 ]</td>
</tr>
<tr>
<td>Jouriles et al. (2016a)</td>
<td>0.33 [ 0.06 , 0.60 ]</td>
</tr>
<tr>
<td>Jouriles et al. (2016b)</td>
<td>0.30 [ 0.01 , 0.59 ]</td>
</tr>
<tr>
<td>Kleinsasser et al. (2015)</td>
<td>0.02 [ -0.39 , 0.43 ]</td>
</tr>
<tr>
<td>Miller et al. (2013)</td>
<td>0.20 [ -0.23 , 0.63 ]</td>
</tr>
<tr>
<td>Moynihan et al. (2010)</td>
<td>0.01 [ -0.40 , 0.42 ]</td>
</tr>
<tr>
<td>Sargent et al. (2017)</td>
<td>0.13 [ -0.09 , 0.35 ]</td>
</tr>
<tr>
<td>Random</td>
<td>0.23 [ 0.13 , 0.33 ]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.27 [ 0.19 , 0.36 ]</td>
</tr>
</tbody>
</table>

We conducted mixed-effects meta-regression to examine whether the effect of bystander programs on bystander intervention behavior varied based on the following: (1) gendered content of the program (i.e., sex of perpetrators & victims); (2) mixed- or single-sex group implementation; (3) gender composition of the sample (i.e., proportion males); (4) education level of the sample (i.e., secondary school or college students); and (5) mean age of the sample. Although we had also planned to examine proportion of fraternity/sorority members in the sample and proportion of athletic team members as potential effect size moderators, these measures were rarely or never reported in the included studies and thus we were unable to examine them. Finally, due to the small number of included studies in this synthesis, we conducted separate bivariate regressions for each potential moderator, meaning that we estimated separate meta-regression models for each potential moderator. As summarized in Table 4, none of these moderators exhibited a statistically significant bivariate association with effect size magnitude. The results from these moderator analyses should be interpreted cautiously, however, given the small number of included studies and our inability to estimate more comprehensive, multivariable meta-regression models.
Table 4: Unstandardized Bivariate Meta-regression Coefficients for Potential Moderators of Program Effects on Bystander Intervention Behavior – One to Four Months Post-intervention

<table>
<thead>
<tr>
<th>Moderator</th>
<th>B</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex of Perpetrators &amp; Victims (Gender Neutral)</td>
<td>0.10</td>
<td>0.17</td>
<td>[-0.23, 0.43]</td>
</tr>
<tr>
<td>Mixed Sex Implementation (Yes)</td>
<td>0.13</td>
<td>0.17</td>
<td>[-0.21, 0.46]</td>
</tr>
<tr>
<td>Proportion Males in Sample</td>
<td>-0.00</td>
<td>0.00</td>
<td>[-0.01, 0.00]</td>
</tr>
<tr>
<td>Education Level (Secondary School)</td>
<td>-0.16</td>
<td>0.11</td>
<td>[-0.38, 0.05]</td>
</tr>
<tr>
<td>Age</td>
<td>0.04</td>
<td>0.02</td>
<td>[-0.01, 0.08]</td>
</tr>
</tbody>
</table>

NOTES: The reference group for Sex of Perpetrators and Victims is male perpetrators/female victims. The reference group for Education Level is college. Coefficients were estimated from separate bivariate regressions for each moderator.

To examine small study and publication bias we created a contour-enhanced funnel plot of the 11 effect sizes plotted against their standard errors (see Figure 32). Visual inspection of the funnel plot reveals an absence of adverse intervention effects. Given the absence of negative effects in the regions of statistical significance and non-significance, the results from this contour-enhanced funnel plot indicate a potential risk of publication bias.

To further investigate the possibility of bias, we conducted an Egger test for funnel plot asymmetry. The results provided no significant evidence of small study effects (bias coefficient: 0.36; t: -0.60, p = .56). Finally, we also conducted a trim and fill analysis. With one study trimmed and filled, the resulting mean effect size was similar to that observed in the main analysis (trim-and-fill mean \( g = 0.28 \), 95% CI [0.20, 0.37]). With these collective findings, we therefore conclude that the meta-analysis results shown in Figure 31 are likely robust to any small study/publication bias.

**Figure 32: Contour Enhanced Funnel Plot of Bystander Intervention 1 to 4 Months Post-Intervention**
We also conducted meta-regression to examine whether the effect of bystander programs on bystander intervention outcomes varied based on the following characteristics: (1) attrition at first follow-up (i.e., one to four months post-intervention), (2) random sequence ROB, (3) blinding of participants ROB, (4) incomplete data ROB, and (5) selective reporting ROB. We had also planned to examine blinding of assessment ROB, but the lack of variation in this measure precluded such analysis (i.e., all studies in this meta-analysis were coded as high ROB for blinding of assessment). Due to the small sample size, we conducted bivariate regressions for each potential moderator.

As shown in Table 5, there was no evidence that attrition, blinding of participants, incomplete data, or selective reporting risk of bias assessments were associated with effect size magnitude. There was, however, evidence that studies rated as unclear risk of bias due to random sequence generation reported significantly smaller effect sizes than those at low risk of bias. There was no evidence that studies rated as high risk of bias due to random sequence generation reported significantly different effect sizes from those at low risk of bias.

Table 5: Unstandardized Bivariate Meta-regression Coefficients for Attrition and Risk of Bias Moderators of the Relationship Between Bystander Programs and Bystander Intervention One- to Four-Months Post-Tx

<table>
<thead>
<tr>
<th>Moderator</th>
<th>B</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attrition at First Follow-Up</td>
<td>-0.39</td>
<td>0.38</td>
<td>[-1.14, 0.35]</td>
</tr>
<tr>
<td>Random Sequence Bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>0.13</td>
<td>0.11</td>
<td>[-0.09, 0.35]</td>
</tr>
<tr>
<td>Unclear</td>
<td>-0.21*</td>
<td>0.10</td>
<td>[-0.41, 0.00]</td>
</tr>
<tr>
<td>Blinding of Participants</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unclear</td>
<td>-0.16</td>
<td>0.11</td>
<td>[-0.37, 0.06]</td>
</tr>
<tr>
<td>Incomplete Data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>0.20*</td>
<td>0.11</td>
<td>[-0.00, 0.41]</td>
</tr>
<tr>
<td>Unclear</td>
<td>-0.03</td>
<td>0.23</td>
<td>[-0.47, 0.41]</td>
</tr>
<tr>
<td>Selective Reporting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>-0.12</td>
<td>0.24</td>
<td>[-0.58, 0.35]</td>
</tr>
</tbody>
</table>

NOTES: *p < .05; †p < .10. Low is the reference group for all risk of bias predictors. For Blinding of Participants no studies were coded as high risk. For Selective Reporting no studies were coded as unclear risk. Coefficients were estimated from separate bivariate regressions for each moderator.

We conducted a sensitivity analysis to determine whether Winsorizing the effect size for Peterson et al. (2016) affected our overall findings. Running the meta-analysis with the original effect size for this study produced identical findings to those from the main analysis for both the two non-randomized studies ($g = 0.43$, 95% CI [0.25, 0.61]) and the total sample containing both randomized and non-randomized studies ($g = 0.27$, 95% CI [0.19, 0.36]). Tests for heterogeneity revealed non-significant heterogeneity, although specific values were slightly different than those from the main analysis for both the non-randomized studies ($X^2 = 0.72$ [p = .40], $I^2 = 0.00\%$, $\tau^2 = 0$) and the total sample ($X^2 = 9.93$ [p = .45], $I^2 = 2.63\%$, $\tau^2 = 0$).
We also conducted two sensitivity analyses in which we replaced (1) Banyard et al.’s (2007) one-session intervention arm with the three-session treatment arm and (2) Jouriles et al.’s (n.d.) independent intervention arm with the monitored intervention arm (i.e., completed in a lab with supervision). As indicated in Figure 33, findings from the Banyard et al. sensitivity analysis were substantively similar to those from the main analysis. The average intervention effect for the nine randomized studies (which included Banyard et al.) was significant and positive ($g = 0.20, 95\% \text{ CI} [0.10, 0.30]$) with non-significant heterogeneity ($X^2 = 3.64 [p = .89], I^2 = 0.00\%, \tau^2 = 0$). The average intervention effect (for randomized and non-randomized studies) was also significant and positive ($g = 0.25, 95\% \text{ CI} [0.16, 0.34]$) with non-significant heterogeneity ($X^2 = 9.10 [p = .52], I^2 = 0.93\%, \tau^2 = 0$).

**Figure 33: Forest Plot of Bystander Intervention Effects on Bystander Intervention – One to Four Months Post-intervention with Alternative Tx Arm for Banyard et al. (2007)**

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges $g$ [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Peterson et al. (2016)</td>
<td>0.60 [0.13, 1.07]</td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td>0.40 [0.20, 0.60]</td>
</tr>
<tr>
<td>Nonrandom</td>
<td>0.43 [0.25, 0.61]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Banyard et al. (2007)</td>
<td>0.17 [-0.08, 0.42]</td>
</tr>
<tr>
<td>Gidycz et al. (2011)</td>
<td>0.16 [-0.29, 0.61]</td>
</tr>
<tr>
<td>Jouriles et al. (n.d.)</td>
<td>0.26 [0.04, 0.48]</td>
</tr>
<tr>
<td>Jouriles et al. (2016a)</td>
<td>0.33 [0.06, 0.60]</td>
</tr>
<tr>
<td>Jouriles et al. (2016b)</td>
<td>0.30 [0.01, 0.59]</td>
</tr>
<tr>
<td>Kleinsasser et al. (2015)</td>
<td>0.02 [-0.39, 0.43]</td>
</tr>
<tr>
<td>Miller et al. (2013)</td>
<td>0.20 [-0.23, 0.63]</td>
</tr>
<tr>
<td>Moynihan et al. (2010)</td>
<td>0.01 [-0.40, 0.42]</td>
</tr>
<tr>
<td>Sargent et al. (2017)</td>
<td>0.13 [-0.09, 0.35]</td>
</tr>
<tr>
<td>Random</td>
<td>0.20 [0.10, 0.30]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.25 [0.16, 0.34]</td>
</tr>
</tbody>
</table>

Similarly, as indicated in Figure 34, findings from the Jouriles et al. (n.d.) sensitivity analysis were substantively similar to those from the main analysis. The average intervention effect for the nine randomized studies (which included Jouriles et al., n.d.) was significant and positive ($g = 0.20, 95\% \text{ CI} [0.11, 0.30]$) with non-significant heterogeneity ($X^2 = 5.60 [p = .69], I^2 = 0.00\%, \tau^2 = 0$). The overall average effect (for randomized and non-randomized studies) was also significant and positive ($g = 0.25, 95\% \text{ CI} [0.16, 0.35]$) with non-significant heterogeneity across individual study effect sizes ($X^2 = 10.80 [p = .37], I^2 = 12.69\%, \tau^2 = 0$).
**Six Month to One-Year Follow-Up.** Four studies reported bystander behavior six months to one year post-intervention. As shown in the forest plot in Figure 35, the program effect for the single non-randomized study in this subsample was non-significant ($g = -0.06, 95\% \text{ CI} [-0.41, 0.29]$) and the average intervention effect for the three randomized studies was non-significant ($g = 0.20, 95\% \text{ CI} [-0.08, 0.32]$) with non-significant heterogeneity ($X^2 = 1.08 [p = .58], I^2 = 0.00\%, \tau^2 = 0$). Across all studies (randomized and non-randomized), the average intervention effect was also non-significant ($g = 0.12, 95\% \text{ CI} [-0.08, 0.32]$) with non-significant heterogeneity ($X^2 = 2.60 [p = .46], I^2 = 14.23\%, \tau^2 = .01$). Thus, these findings indicate that, at six months to one year post-intervention, bystander programs have a significant positive (beneficial) effect on bystander intervention. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.
Figure 35: Forest Plot of Bystander Intervention Effects on Bystander Intervention - Six Months to One Year Post-intervention

<table>
<thead>
<tr>
<th>Study</th>
<th>Hedges g [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nonrandom Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Miller et al. (2014)</td>
<td>-0.06 [-0.41, 0.29]</td>
</tr>
<tr>
<td><strong>Random Assignment</strong></td>
<td></td>
</tr>
<tr>
<td>Gidycz et al. (2011)</td>
<td>0.07 [-0.38, 0.52]</td>
</tr>
<tr>
<td>Miller et al. (2013)</td>
<td>0.04 [-0.49, 0.57]</td>
</tr>
<tr>
<td>Salazar et al. (2014)</td>
<td>0.29 [0.02, 0.56]</td>
</tr>
<tr>
<td>Random</td>
<td>0.20 [-0.02, 0.41]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>0.12 [-0.08, 0.32]</td>
</tr>
</tbody>
</table>

**Sexual assault perpetration**
Six studies reported sexual assault perpetration as an outcome, asking participants to self-report their perpetration of sexual coercion or sexual abuse. All but one of these studies (Miller et al., 2014) involved random assignment of participants to conditions. All six studies were published in a peer-reviewed outlet.

No studies measured sexual assault perpetration at immediate post-intervention. We thus collapsed bystander intervention effect sizes into two timing intervals: (1) three to four months post-intervention and (2) six months to one year post-intervention.

**Three to Four-Month Follow-Up.** Two studies measured sexual assault perpetration three or four months post-intervention. As shown in the forest plot in Figure 36, the average intervention effect across these two studies was non-significant ($g = 0.33$, 95% CI [-0.70, 1.36]) with non-significant heterogeneity across individual study effect sizes ($X^2 = 0.45 [p = .50]$, $I^2 = 0.00\%$, $\tau^2 = 0$). These findings provide no evidence that bystander programs have an effect on sexual assault perpetration at three- to four-months post-intervention.
Six Month to One-Year Follow-up. Four studies measured sexual assault perpetration six months to one year post-intervention. As shown in the forest plot in Figure 37, the program effect for the one non-randomized study was non-significant ($g = 0.11, 95\% \text{ CI [-0.75, 0.97]}$). The average intervention effect across the three randomized studies was non-significant ($g = 0.10, 95\% \text{ CI [-0.12, 0.32]}$) with non-significant heterogeneity ($X^2 = 2.56 [p = .28], I^2 = 35.66\%, \tau^2 = 0.01$). Across all studies (randomized and non-randomized), the average intervention was non-significant ($g = 0.10, 95\% \text{ CI [-0.10, 0.30]}$) with non-significant heterogeneity ($X^2 = 2.57 [p = .46], I^2 = 23.85\%, \tau^2 = 0.01$). These findings provide no evidence that bystander programs have an effect on sexual assault perpetration six months to one year post-intervention. The small sample size in this meta-analysis ($n < 10$) precluded ad-hoc analysis of moderators or small study/publication bias.

Figure 37: Forest Plot of Bystander Intervention Effects on Sexual Assault Perpetration – Six Months to One-Year Post-intervention
Discussion

Summary of main results

Objective 1: Bystander intervention effects on knowledge, attitudes, and behavior

Knowledge/Attitudes. Bystander intervention effects on knowledge/attitude outcomes varied widely across constructs. The most pronounced effect in this domain was on rape myth acceptance. The average program effect for this outcome was immediate and sustained across all reported follow-up waves (i.e., from immediate post-test to six- to seven-months post-intervention). At each of these follow-up waves rape myth acceptance scores were approximately two-fifths of a standard deviation lower for young people who participated in a bystander program relative to scores among the comparison group. To put this in context, one study in our meta-analytic sample (Salazar et al., 2014) reported a pre-intervention mean score of 36.69 (SD = 10.29) on the Rape Myth Acceptance Scale, with a total possible score of 85 and higher scores indicating greater rape myth endorsement. Considering that the average effect size for rape myth acceptance at immediate post-test in our sample was 0.37, extrapolation suggests that participating in a bystander program decreased the mean rape myth acceptance score from 36.69 to 32.88 (i.e., (36.69 - [0.37 * 10.29]) = 32.88), a desirable change of approximately 4 points on an 85-point scale.

Program effects on bystander efficacy were also fairly pronounced, with a significant effect observed at both immediate post-test and one- to four-months post-intervention. At both of these follow-up waves measures of bystander efficacy were approximately one-half of a standard deviation higher for bystander program participants than for the comparison group. Again, for illustrative purposes, one study in our meta-analytic sample (Senn & Forest, 2016) reported a pre-intervention mean score of 62.47 (SD = 20.58) on the Bystander Efficacy Scale, with a total possible score of 100 and lower scores representing greater bystander efficacy. Considering that the average effect size for bystander efficacy at immediate post-test in our sample was 0.45, extrapolation suggests that participating in a bystander program decreased the mean bystander efficacy score from 62.47 to 53.21 (i.e., (62.47 - [0.45 * 20.58]) = 53.21), a desirable change of approximately 9 points on a 100-point scale.

Program effects on identifying a situation as appropriate for intervention were significant but were smaller at later follow-up waves. At immediate post-test measures of this outcome were almost six-tenths of a standard deviation higher for bystander program participants than for the comparison group. By one- to four-months follow-up this difference was four-tenths. Only one study measured
identifying a situation as appropriate for intervention past six months follow-up. To put the immediate post-test findings into context, one study in our meta-analytic sample (Addison, 2015) reported a pre-intervention mean rating of 5.94 ($SD = 1.11$) regarding responses to a vignette depicting a scenario in which a young woman was too inebriated to consent to sexual activity. The total possible score was 7, with higher scores indicating stronger agreement that the scenario depicted in the vignette warranted intervention. Considering that the average effect size for this outcome was 0.57 at immediate post-test, extrapolation suggests that participating in a bystander program increased the mean rating from 5.94 to 6.57 (i.e., $5.94 + [0.57 \times 1.11] = 6.57$), a change of approximately two-thirds a point on a 7-point scale.

For three of the attitude/knowledge outcomes, effects were non-significant at immediate post-test, but significant at the one- to four-month post-intervention follow-up period. This was the case for taking responsibility to intervene, knowing strategies for intervening, and intentions to intervene. At one-to-four months follow-up, relative to the comparison group, bystander program participants reported greater responsibility to intervene (difference of approximately one-third a standard deviation). To put these findings into context, one study in our meta-analytic sample (Senn & Forest 2016) reported a pre-intervention mean score of 3.51 ($SD = 1.13$) on a measure of responsibility to intervene, with a total possible score of 7 and lower scores indicating greater responsibility. Considering that the average effect size for this outcome was 0.32, extrapolation suggests that participating in a bystander program decreased the mean score on this outcome from 3.51 to 3.15 (i.e., $3.51 + [0.32 \times 1.13] = 3.15$), a desirable change of approximately four-tenths a point on a 7-point scale.

At one-to-four months follow-up, relative to the comparison group, bystander program participants reported higher measures of knowing strategies for intervening (approximately six-tenths of a standard deviation higher). However, this effect was only based on one study.

At one-to-four months follow-up, relative to the comparison group, bystander program participants reported higher measures of intentions to intervene (approximately four-tenths a standard deviation higher). To put these findings into context, one study in our meta-analytic sample (Salazar et al., 2014) reported a pre-intervention mean score of 52.09 ($SD = 12.78$) on the Bystander Intention Scale, with a total possible score of 75 and higher scores indicating greater intentions. Considering that the average effect size for this outcome was 0.41 at the one-to-four months follow-up wave, extrapolation suggests that participating in a bystander program increased the mean score on this outcome from 52.09 to 57.33 (i.e., $52.09 + [0.41 \times 12.78] = 57.33$), a desirable change of approximately 5 points on a 75-point scale. Significant effects for this outcome were also observed at the six months to one-year follow-up wave (approximately one-fourth a standard deviation). Considering the effect size at this follow-up wave was 0.23, participation in a bystander program increased the mean bystander intention score from 52.09 to 55.03 between pre-intervention and the six-months to one year follow-up (i.e., $52.09 + [0.23 \times 12.78] = 55.03$), a change of approximately 3 points on a 75-point scale. Additionally, the mean score on bystander intentions decreased from 57.33 to 55.03 between one-to-four months follow-up and six months to one-year follow-up.
The fact that significant effects were delayed until one-to-four months follow-up for three attitude/knowledge outcomes may indicate that participants’ knowledge and attitudes shifted over time or that the programs required a bit of rumination and reflection to take effect. We found limited or no evidence of intervention effects on gender attitudes, victim empathy, date rape attitudes, and noticing sexual assault.

**Behavior.** Our results indicated that bystander programs have a desirable effect on bystander intervention. At one- to four-months follow-up measures of bystander intervention were approximately one-fourth a standard deviation higher for bystander program participants than for the comparison group. To contextualize, one study in our meta-analytic sample (Jouriles et al., 2016a) reported a pre-intervention mean score of 27.95 (SD = 19.02) on the Bystander Intervention Scale, with a total possible score of 49 and higher scores indicating greater intervention behavior over the past month. Considering that the average effect size for bystander intervention at one- to four-months follow-up was 0.27, then extrapolation suggests that participation in a bystander intervention program increased bystander intervention scores from 27.95 to 33.09 (i.e., \(27.95 + (0.27 \times 19.02)\) = 33.09), indicating that participants engaged in approximately five additional acts of intervention in the past month (relative to pre-test behavior). However, this effect was non-significant at six months post-intervention.

It is noteworthy that bystander programs did not have a significant effect on sexual assault perpetration. This finding seems to belie the original intent of some of the earliest bystander programs, which aimed to reduce sexual violence perpetration by approaching young people (young men especially) as allies in preventing violence against women, rather than as potential perpetrators of violence. This non-threatening approach was anticipated to be more effective at reducing violence against women than traditional victim/perpetrator-focused models, which may risk young men’s defensiveness or backlash (Messner, 2015; Katz, 1995). Instead, evidence from our review indicates that, although bystander programs may have a meaningful, short-term effect on bystander intervention behavior, there is no evidence that the programs have a similar effect on sexual assault perpetration. Thus, it appears that bystander program participants interpret program information from the perspective of potential allies rather than potential perpetrators.

**Objective 2: Effects for different participant profiles**

We had planned to conduct moderator analyses to assess any differential effects of bystander programs on outcomes based on the following participant characteristics: mean age of the sample, education level of the sample, proportion of males/females in the sample, proportion of fraternity/sorority members in the sample, and proportion of athletic team members in the sample. Our review only produced a sufficient number of studies (\(n \geq 10\)) to conduct such moderator analyses for the bystander intervention behavior outcome. Those results provided no evidence that mean age, education level, or proportion of males/female were associated with the magnitude of program effects on bystander intervention behavior. We were unable to conduct moderator analyses for proportion of fraternity/sorority members in the sample and proportion of athletes in the sample, as these variables were rarely or never reported.
Objective 3: Effects based on gendered content/implementation

We conducted moderator analyses to assess any differential effects of bystander programs on measured outcomes based on (1) the gender of perpetrators and victims in specific bystander programs and (2) whether programs were implemented in mixed- or single-sex settings. Our review only produced a sufficient number of studies \( (n > 10) \) to conduct such moderator analyses for the bystander intervention behavior outcome. The results provided no evidence that these measures were associated with the magnitude of program effects on bystander intervention behavior.

Overall completeness and applicability of evidence

The thorough nature of our systematic literature search process (e.g., searching electronic databases, ClinicalTrials.gov, conference proceedings, organization websites, reference lists of review articles and eligible reports, tables of contents of relevant journals, CVs and websites of primary authors of eligible studies, and forward citation searching of eligible studies) helped ensure that all relevant studies were identified for consideration in this review. Nonetheless, there were three rogue reports that we were unable to locate. Readers should be mindful of this when interpreting findings from this review.

Readers should also be mindful that, because bystander sexual assault prevention programs are not standardized, the summarized research employed a variety of modalities, program curricula, and delivery methods. In order to be included in this systematic review and meta-analysis, all programs had to include a bystander intervention component, but there were no other restrictions or limitations on programs. Readers should therefore consider that our syntheses include different types of programs. Although this does not necessarily limit the applicability or generalizability of the systematic review and meta-analysis, readers should be cognizant that a variety of bystander intervention programs may be included in each quantitative synthesis.

Although our systematic literature search identified 27 eligible studies, only 13 reported bystander behavior and only six reported sexual assault. Thus, results for these behavioral outcomes may be considered tentative, as an updated review that includes more studies may produce different findings. Additionally, included studies rarely reported some of the moderators that we identified in our protocol (i.e., fraternity/sorority membership and athletic team membership) and the vast majority (i.e., 25 of 27) were conducted in the United States. Thus, we do not know how well fraternity/sorority members and athletic team members were represented in the sample and we do know that participants from the United States were over-represented.

This paucity of international research highlights a clear need for further study of bystander programs in a global context and serves to underscore the necessity of forming a broader evidence base. Readers should be cautioned that the findings of this review may not generalize to all contexts, especially those that are demographically different from the United States.

Finally, no eligible studies examined sexual assault in lesbian, gay, bisexual, or queer relationships or interactions, and programs reported in the literature did not make clear that they mentioned or
addressed potential sexual assault in non-heterosexual relationships or interactions. Further, eligible studies did not indicate that they provided any particular or tailored strategies for being a bystander in non-heterosexual situations.

**Quality of the evidence**

Although we deemed a large number of studies to be ineligible based on methodological quality, the large body of extant research on bystander program permitted the identification of 27 independent studies that met all inclusion criteria for this review. Of these 27 studies, 21 were RCTs (i.e., 12 randomized at the individual level and 9 randomized at the group level) and 6 were high-quality quasi-experimental studies.

Risk of bias for the RCTs in this review was relatively low; however, there were some issues that must be noted. First, we coded approximately 95% of the sample of RCTs as exhibiting high risk of bias pertaining to blinding of outcome assessments. This was due to the frequent reporting of self-report outcomes. Second, we coded approximately 80% of the sample of RCTs as exhibiting high risk of bias in some “other” domain. Such designation was typically the result of study authors evaluating programs that they designed themselves. Risk of bias for the nine non-RCT studies was typically low or unclear. The latter designation was typically a result of study authors’ failure to report our pre-specified confounding variables broken down by treatment and comparison groups.

**Limitations and potential biases in the review process**

As a whole, the existing evidence base for this review (N = 27) was fairly strong methodologically, but the relatively small number of eligible studies in each meta-analysis precluded our analysis of important sources of heterogeneity. That is, only one meta-analysis contained sufficient studies (n ≥ 10) to conduct moderator analyses and explore issues stemming from risk of bias or publication/small study bias. Future accumulation of high-quality research should permit such analyses in updated reviews. Additionally, the dearth of non-US research is a limitation to the generalizability of the findings.

**Agreements and disagreements with other studies or reviews**

To date, there is only one existing meta-analysis examining the effects of bystander programs. In what they called an “initial” meta-analysis of experimental and quasi-experimental studies (published through 2011), Katz & Moore (2013) found moderate effects of bystander programs on participants’ self-efficacy and intentions to intervene, and small (but significant) effects on bystander behavior, rape-supportive attitudes, and rape proclivity (but not perpetration). Effects were generally stronger among younger samples and samples containing a higher percentage of males.

The main findings from the current review are consistent with those reported by Katz and Moore. Specifically, like Katz and Moore’s analysis, our review findings indicated that bystander programs
have beneficial effects on bystander efficacy and intentions. Additionally, we found beneficial effects on bystander behavior (i.e., bystander intervention) and rape supportive attitudes (i.e., rape myth acceptance). Unlike Katz and Moore, however, we did not analyze rape proclivity outcomes; but similar to Katz and Moore, we found no evidence of an effect on sexual assault perpetration.

Katz and Moore also conducted moderator analyses to assess differential effects of bystander programs on (1) bystander efficacy and (2) bystander intentions. They reported that program effects on bystander efficacy were stronger among younger participants and that effects on bystander intentions were stronger among samples containing a higher proportion of males. Our moderator analyses provided no evidence that participant age and proportion of males were predictors of program effects. However, we were unable to explore these moderators in depth given the small number of studies included in each meta-analysis.

Although findings from our meta-analysis are largely similar to those of Katz and Moore, our analysis advances our understanding of the evidence base for bystander programs in two important ways. First, whereas Katz and Moore’s analysis focused on bystander programs implemented with college students, our analysis focused on programs implemented with both college students and adolescents. Given that our moderator analyses indicated age and education level were not significant moderators of program effects on bystander intervention, our findings for this outcome may be representative of both college students and adolescents. However, findings from these moderator analyses should be interpreted as preliminary, given the small number of studies reporting bystander intervention.

Second, whereas Katz and Moore did not evaluate program content/implementation as a moderator of program effects, we were able to examine the influence of such variables on an important outcome. We found that (1) sex of perpetrators/victims in bystander programs and (2) whether programs were implemented in mixed- or single-group settings were not significant moderators of program effects on bystander intervention. Again, however, findings from these moderator analyses should be interpreted as preliminary given the small number of studies reporting bystander intervention.
Authors’ conclusions

Implications for practice and policy

The United States 2013 Campus Sexual Violence Elimination (SaVE) Act requires post-secondary educational institutions participating in Title IX financial aid programs to provide incoming college students with primary prevention and awareness programs addressing sexual violence. The Campus SaVE Act mandates that these programs include a component on bystander intervention. Currently, there is no comparable legislation regarding sexual assault among adolescents (e.g., mandating bystander programs in secondary schools). This is an unfortunate oversight, as adolescents who experience sexual assault are at an increased risk of repeated victimization in young adulthood (Cui et al., 2013). Thus, the implementation of bystander programs in secondary schools not only has the potential to reduce sexual assault among adolescents, but may also have the long-term potential to reduce sexual assault on college campuses.

Findings from this review indicate that bystander programs have significant beneficial effects on bystander intervention behavior. This provides important evidence of the effectiveness of mandated programs on college campuses. Additionally, the fact that our (preliminary) moderator analyses found program effects on bystander intervention to be similar for adolescents and college students suggests early implementation of bystander programs (i.e., in secondary schools with adolescents) may be warranted.

Importantly, although we found that bystander programs had a significant beneficial effect on bystander intervention behavior, we found no evidence that these programs had an effect on participants’ sexual assault perpetration. Bystander programs may therefore be appropriate for targeting bystander behavior, but may not be appropriate for targeting the behavior of potential perpetrators. Additionally, effects of bystander programs on bystander intervention behavior diminished by six-months post-intervention. Thus, programs effects may be prolonged by the implementation of booster sessions conducted prior to six months post-intervention.

Implications for research

Findings from this review suggest there is a fairly strong body of research assessing the effects of bystander programs on attitudes and behaviors. However, there are a couple of important questions worth further exploration.
First, according to one prominent logical model, bystander programs promote bystander intervention by fostering prerequisite knowledge and attitudes (Burn, 2009). Our meta-analysis provides inconsistent evidence of the effects of bystander programs on knowledge and attitudes, but promising evidence of short-term effects on bystander intervention. This casts uncertainty around the proposed relationship between knowledge/attitudes and bystander behavior. Although we were unable to assess these issues in the current review, this will be an important direction for future research. Our understanding of the causal mechanisms of program effects on bystander behavior would benefit from further analysis (e.g., path analysis mapping relationships between specific knowledge/attitude effects and bystander intervention).

Second, bystander programs exhibit a great deal of content variability, most notably in framing sexual assault as a gendered or gender-neutral problem. That is, bystander programs tend to adopt one of two main approaches to addressing sexual assault: (1) presenting sexual assault as a gendered problem (overwhelmingly affecting women) or (2) presenting sexual assault as a gender-neutral problem (affecting women and men alike). Differential effects of these two types of programs remain largely unexamined. Our analysis indicated that (1) the sex of victims/perpetrators (i.e., portrayed in programs as gender neutral or male perpetrator and female victim) and (2) whether programs were implemented in mixed- or single-sex settings were not significant moderators of program effects on bystander intervention. However, these findings are limited to a single outcome and they should be considered preliminary, as they are based on a small sample ($n = 11$). Our understanding of the differential effects of gendered versus gender neutral programs would benefit from the design and implementation of high-quality primary studies that make direct comparisons between these two types of programs (e.g., RCTs comparing the effects of two active treatment arms that differ in their gendered approach).

Finally, our systematic review and meta-analysis demonstrates the lack of global evidence concerning bystander program effectiveness. Our understanding of bystander programs’ generalizability to non-US contexts would be greatly enhanced by high quality research conducted across the world.
References to included studies


Jouriles, E. et al. (n.d.). TakeCARE, a Video to Promote Bystander Behavior on College Campuses: Replication and Extension. Unpublished manuscript.


References to excluded studies


References to studies awaiting classification


References to ongoing studies


Griffin, K. W. Primary Prevention of Sexual Violence Among College Students. ClinicalTrials.gov Identifier: NCT03037866.

Jouries, E. Evaluating a Video Bystander Program for First-Year College Students. ClinicalTrials.gov Identifier: NCT03056560.

Miller, E. College Health Center-based Alcohol and Sexual Violence Intervention (GIFTSS). ClinicalTrials.gov Identifier: NCT02355470.
Additional references


Information about this review

Review authors

Lead review author

<table>
<thead>
<tr>
<th>Name:</th>
<th>Heather Hensman Kettrey, PhD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Assistant Professor</td>
</tr>
<tr>
<td>Affiliation:</td>
<td>Department of Sociology, Anthropology, and Criminal Justice Clemson University</td>
</tr>
<tr>
<td>Address:</td>
<td>132 Brackett Hall</td>
</tr>
<tr>
<td>City, State, Province or County:</td>
<td>Clemson, SC</td>
</tr>
<tr>
<td>Post code:</td>
<td>29634-1356</td>
</tr>
<tr>
<td>Country:</td>
<td>USA</td>
</tr>
<tr>
<td>Phone:</td>
<td>(864) 656-1107</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:hkettre@clemson.edu">hkettre@clemson.edu</a></td>
</tr>
</tbody>
</table>

Co-authors

<table>
<thead>
<tr>
<th>Name:</th>
<th>Robert A. Marx, MA, MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Doctoral Candidate</td>
</tr>
<tr>
<td>Affiliation:</td>
<td>Department of Human and Organizational Development Peabody College Vanderbilt University</td>
</tr>
<tr>
<td>Address:</td>
<td>230 Appleton Place</td>
</tr>
<tr>
<td>City, State, Province or County:</td>
<td>Nashville, TN</td>
</tr>
<tr>
<td>Post code:</td>
<td>37203-5721</td>
</tr>
<tr>
<td>Country:</td>
<td>USA</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:Robert.a.marx@vanderbilt.edu">Robert.a.marx@vanderbilt.edu</a></td>
</tr>
</tbody>
</table>
Roles and responsibilities

Members of the research team responsible for core areas of the review are as follows:

- **Content**: Heather Hensman Kettrey
- **Systematic review methods**: Heather Hensman Kettrey, Robert A. Marx, Emily E. Tanner-Smith
- **Statistical analysis**: Heather Hensman Kettrey
- **Information retrieval**: Heather Hensman Kettrey, Robert A Marx

Dr Kettrey, the lead review author, coordinated the review team and assumed responsibility for the implementation of the project throughout its duration. Specific tasks included compiling the sample of research reports, creating the database, coding studies, analyzing data, and preparing the Campbell Review. Mr Marx, the second review author, collaborated closely with Dr Kettrey to compile the sample of research reports, code studies, and make methodological decisions throughout the duration of the project. Dr Tanner-Smith, the third review author, provided methodological guidance and mentorship to Dr Kettrey and Mr Marx throughout all phases of data collection and analysis.

Sources of support

This review was supported by a grant from Campbell (CSR1.60).

Declarations of interest

The authors have no conflicts of interests to report.
Plans for updating the review

Heather Hensman Kettrey anticipates updating the review at least once every five years, pending continual accrual of sufficient research on the topic. If she is unable to assume this responsibility she will transfer responsibility to another qualified researcher – with approval from Campbell.
### Figures and tables

**Table A1: Characteristics of included studies**

<table>
<thead>
<tr>
<th>Study Primary Report</th>
<th>Participants</th>
<th>Methods</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Addison (2015)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational Setting: College/University</td>
<td>Male: 53.5%</td>
<td>Design: RCT</td>
<td>Random Sequence: Low</td>
</tr>
<tr>
<td>Male: 53.5%</td>
<td>White: 40.7%</td>
<td>Funded: No</td>
<td>Allocation: Low</td>
</tr>
<tr>
<td>White: 40.7%</td>
<td>Country: United States</td>
<td>Peer Reviewed: No</td>
<td>Blinding of Participants: Unclear</td>
</tr>
<tr>
<td>Assigned Treatment N: 164</td>
<td>Primary Report Type: Dissertation</td>
<td>Random Sequence: High</td>
<td>Blinding of Assessment: High</td>
</tr>
<tr>
<td>Assigned Comparison N: 150</td>
<td>Treatment Program: Sexual Assault Bystander Awareness</td>
<td>Allocation: Unclear</td>
<td>Incomplete Data: Low</td>
</tr>
<tr>
<td></td>
<td>Comparison Type: Inactive Comparison</td>
<td>Selective Reporting: Low</td>
<td>Other: High</td>
</tr>
<tr>
<td></td>
<td>Other: Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Amar et al. (2015)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational Setting: College/University</td>
<td>Male: 0%</td>
<td>Design: QED</td>
<td>Random Sequence: High</td>
</tr>
<tr>
<td>Male: 0%</td>
<td>White: 85.3%</td>
<td>Funded: Yes</td>
<td>Allocation: Unclear</td>
</tr>
<tr>
<td>White: 85.3%</td>
<td>Country: United States</td>
<td>Peer Reviewed: Yes</td>
<td>Blinding of Participants: Unclear</td>
</tr>
<tr>
<td>Assigned Treatment N: 40</td>
<td>Primary Report Type: Journal Article</td>
<td>Random Sequence: High</td>
<td>Blinding of Assessment: High</td>
</tr>
<tr>
<td>Assigned Comparison N: 64</td>
<td>Treatment Program: Friends Helping Friends</td>
<td>Allocation: Unclear</td>
<td>Incomplete Data: Unclear</td>
</tr>
<tr>
<td></td>
<td>Comparison Type: Inactive Comparison</td>
<td>Selective Reporting: High</td>
<td>Other: Low</td>
</tr>
<tr>
<td></td>
<td>Other: Low</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Primary Report</td>
<td>Participants</td>
<td>Methods</td>
<td>Risk of Bias</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------</td>
<td>---------</td>
<td>--------------</td>
</tr>
</tbody>
</table>
| **Baker et al. (2014)** | **Educational Setting:** Secondary School  
**Male:** 44.3%  
**White:** 15.7%  
**Country:** United States  
**Assigned Treatment N:** 65  
**Assigned Comparison N:** 104 | **Design:** QED  
**Funded:** Yes  
**Peer Reviewed:** Yes  
**Primary Report Type:** Journal Article  
**Treatment Program:** Respect  
**Comparison Type:** Inactive Comparison | **Random Sequence:** High  
**Allocation:** Unclear  
**Blinding of Participants:** Unclear  
**Blinding of Assessment:** High  
**Incomplete Data:** High  
**Selective Reporting:** Low  
**Other:** High |
| **Banyard et al. (2007)** | **Educational Setting:** College/University  
**Treatment 1 Male:** 43.9%  
**Treatment 2 Male:** 44.9%  
**White:** 90.4%  
**Country:** United States  
**Assigned Treatment 1 N:** 129  
**Assigned Treatment 2 N:** 124  
**Assigned Comparison N:** 110 | **Design:** RCT  
**Funded:** Yes  
**Peer Reviewed:** Yes  
**Primary Report Type:** Journal Article  
**Treatment Program:** Bringing in the Bystander  
**Comparison Type:** Inactive Comparison | **Random Sequence:** Low  
**Allocation:** Low  
**Blinding of Participants:** Unclear  
**Blinding of Assessment:** High  
**Incomplete Data:** Low  
**Selective Reporting:** Low  
**Other:** High |
| **Brokenshire (2015)** | **Educational Setting:** College/University  
**Male:** Not Reported  
**White:** Not Reported  
**Country:** United States  
**Assigned Treatment N:** 585  
**Assigned Comparison N:** 585 | **Design:** RCT  
**Funded:** No  
**Peer Reviewed:** No  
**Primary Report Type:** Masters Thesis  
**Treatment Program:** Take a Stand!  
**Comparison Type:** Active/Sham | **Random Sequence:** Unclear  
**Allocation:** Unclear  
**Blinding of Participants:** Unclear  
**Blinding of Assessment:** High  
**Incomplete Data:** High  
**Selective Reporting:** Low  
**Other:** High |
| **Chiriboga (2016)** | **Educational Setting:** College/University  
**Male:** 29% | **Design:** RCT  
**Funded:** No | **Random Sequence:** Low  
**Allocation:** Unclear |
<table>
<thead>
<tr>
<th>Study Primary Report</th>
<th>Participants</th>
<th>Methods</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>White:</strong> 74.6%</td>
<td><strong>Country:</strong> United States</td>
<td><strong>Assigned Treatment N:</strong> 78</td>
<td><strong>Assigned Comparison N:</strong> 77</td>
</tr>
<tr>
<td><strong>Country:</strong> United States</td>
<td><strong>Methods:</strong> <strong>Primary Report Type:</strong> Masters Thesis</td>
<td><strong>Methods:</strong> <strong>Treatment Program:</strong> Text Messages</td>
<td><strong>Risk of Bias:</strong> <strong>Comparison Type:</strong> Inactive Comparison</td>
</tr>
<tr>
<td><strong>Blinding of Participants:</strong> Unclear</td>
<td><strong>Blinding of Assessment:</strong> High</td>
<td><strong>Incomplete Data:</strong> High</td>
<td><strong>Selective Reporting:</strong> Low</td>
</tr>
<tr>
<td><strong>Other:</strong> Unclear</td>
<td><strong>Other:</strong> Low</td>
<td><strong>Other:</strong> Low</td>
<td><strong>Other:</strong> Unclear</td>
</tr>
<tr>
<td><strong>Country:</strong> United States</td>
<td><strong>Methods:</strong> <strong>Design:</strong> RCT</td>
<td><strong>Methods:</strong> <strong>Funded:</strong> Yes</td>
<td><strong>Methods:</strong> <strong>Peer Reviewed:</strong> Yes</td>
</tr>
<tr>
<td><strong>Male:</strong> Not Reported</td>
<td><strong>Methods:</strong> <strong>Primary Report Type:</strong> Journal Article</td>
<td><strong>Methods:</strong> <strong>Treatment Program:</strong> Green Dot</td>
<td><strong>Methods:</strong> <strong>Comparison Type:</strong> Waitlist Control</td>
</tr>
<tr>
<td><strong>White:</strong> Not Reported</td>
<td><strong>Methods:</strong> <strong>Random Sequence:</strong> Unclear</td>
<td><strong>Methods:</strong> <strong>Allocation:</strong> Unclear</td>
<td><strong>Methods:</strong> <strong>Blinding of Participants:</strong> Unclear</td>
</tr>
<tr>
<td><strong>Country:</strong> United States</td>
<td><strong>Methods:</strong> <strong>Blinding of Assessment:</strong> High</td>
<td><strong>Methods:</strong> <strong>Incomplete Data:</strong> High</td>
<td><strong>Methods:</strong> <strong>Selective Reporting:</strong> Low</td>
</tr>
<tr>
<td><strong>Assigned Treatment N:</strong> Not Reported</td>
<td><strong>Methods:</strong> <strong>Other:</strong> Low</td>
<td><strong>Methods:</strong> <strong>Other:</strong> Unclear</td>
<td><strong>Methods:</strong> <strong>Other:</strong> Low</td>
</tr>
<tr>
<td><strong>Assigned Comparison N:</strong> Not Reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Primary Report</td>
<td>Participants</td>
<td>Methods</td>
<td>Risk of Bias</td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Assigned Treatment N: 77</td>
<td>How to Help a Sexual Assault Survivor</td>
<td>Selective Reporting: Low</td>
</tr>
<tr>
<td></td>
<td>Assigned Comparison N: 58</td>
<td></td>
<td>Other: High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comparison Type: Inactive Comparison</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male: 100%</td>
<td>Funded: No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>White: 96.4%</td>
<td>Peer Reviewed: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Country: United States</td>
<td>Primary Report Type: Journal Article</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assigned Treatment N: 76</td>
<td>Treatment Program: How to Help a Sexual Assault Survivor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assigned Comparison N: 38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male: 100%</td>
<td>Funded: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>White: 88%</td>
<td>Peer Reviewed: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Country: United States</td>
<td>Primary Report Type: Journal Article</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assigned Treatment N: 109</td>
<td>Treatment Program: How to Help a Sexual Assault Survivor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assigned Comparison N: 46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foubert et al. (2007)</td>
<td>Educational Setting: College/University</td>
<td>Design: RCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Male: 100%</td>
<td>Funded: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>White: 85%</td>
<td>Peer Reviewed: Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Country: United States</td>
<td>Primary Report Type: Journal Article</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assigned Treatment N: 286</td>
<td>Treatment Program: The Men’s Program</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assigned Comparison N: 287</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

103 The Campbell Collaboration | www.campbellcollaboration.org
<table>
<thead>
<tr>
<th>Study Primary Report</th>
<th>Participants</th>
<th>Methods</th>
<th>Risk of Bias</th>
</tr>
</thead>
</table>
| Gidycz et al. (2011) | **Educational Setting:** College/University  
**Male:** 100%  
**White:** Not Reported  
**Country:** United States  
**Assigned Treatment N:** 206  
**Assigned Comparison N:** 254 | **Design:** RCT  
**Funded:** Yes  
**Peer Reviewed:** Yes  
**Primary Report Type:** Journal Article  
**Treatment Program:** The Men’s Program  
**Comparison Type:** Waitlist Control | **Random Sequence:** Unclear  
**Allocation:** Unclear  
**Blinding of Participants:** Unclear  
**Blinding of Assessment:** High  
**Incomplete Data:** Low  
**Selective Reporting:** High  
**Other:** High |
| Jouriles et al. (n.d.) | **Educational Setting:** College/University  
**Male:** 22.6%  
**White:** 59.6%  
**Country:** United States  
**Assigned Treatment 1 N:** 188  
**Assigned Treatment 2 N:** 177  
**Assigned Comparison N:** 192 | **Design:** RCT  
**Funded:** Unclear  
**Peer Reviewed:** No  
**Primary Report Type:** Unpublished Manuscript  
**Treatment Program:** TakeCARE  
**Comparison Type:** Active/Sham | **Random Sequence:** Low  
**Allocation:** Unclear  
**Blinding of Participants:** Unclear  
**Blinding of Assessment:** High  
**Incomplete Data:** Low  
**Selective Reporting:** Low  
**Other:** High |
| Jouriles et al. (2016a) | **Educational Setting:** College/University  
**Male:** 19.6%  
**White:** 84.0%  
**Country:** United States  
**Assigned Treatment N:** 111  
**Assigned Comparison N:** 102 | **Design:** RCT  
**Funded:** Yes  
**Peer Reviewed:** Yes  
**Primary Report Type:** Journal Article  
**Treatment Program:** TakeCARE  
**Comparison Type:** Active/Sham | **Random Sequence:** Low  
**Allocation:** Unclear  
**Blinding of Participants:** Unclear  
**Blinding of Assessment:** High  
**Incomplete Data:** Low  
**Selective Reporting:** Low  
**Other:** High |
<table>
<thead>
<tr>
<th>Study Primary Report</th>
<th>Participants</th>
<th>Methods</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jouriles et al. (2016b)</td>
<td><strong>Educational Setting</strong>: College/University</td>
<td><strong>Design</strong>: RCT</td>
<td><strong>Random Sequence</strong>: Low</td>
</tr>
<tr>
<td></td>
<td><strong>Male</strong>: 50.0%</td>
<td><strong>Funded</strong>: Yes</td>
<td><strong>Allocation</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>White</strong>: 68.5%</td>
<td><strong>Peer Reviewed</strong>: Yes</td>
<td><strong>Blinding of Participants</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>Country</strong>: United States</td>
<td><strong>Primary Report Type</strong>: Journal Article</td>
<td><strong>Blinding of Assessment</strong>: High</td>
</tr>
<tr>
<td></td>
<td><strong>Assigned Treatment N</strong>: 108</td>
<td><strong>Treatment Program</strong>: TakeCARE</td>
<td><strong>Incomplete Data</strong>: Low</td>
</tr>
<tr>
<td></td>
<td><strong>Assigned Comparison N</strong>: 103</td>
<td><strong>Comparison Type</strong>: Active/Sham</td>
<td><strong>Selective Reporting</strong>: Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Other</strong>: High</td>
</tr>
<tr>
<td>Kleinsasser et al. (2015)</td>
<td><strong>Educational Setting</strong>: College/University</td>
<td><strong>Design</strong>: RCT</td>
<td><strong>Random Sequence</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>Male</strong>: 19.4%</td>
<td><strong>Funded</strong>: Yes</td>
<td><strong>Allocation</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>White</strong>: 66.7%</td>
<td><strong>Peer Reviewed</strong>: Yes</td>
<td><strong>Blinding of Participants</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>Country</strong>: United States</td>
<td><strong>Primary Report Type</strong>: Journal Article</td>
<td><strong>Blinding of Assessment</strong>: High</td>
</tr>
<tr>
<td></td>
<td><strong>Assigned Treatment N</strong>: 45</td>
<td><strong>Treatment Program</strong>: TakeCARE</td>
<td><strong>Incomplete Data</strong>: Low</td>
</tr>
<tr>
<td></td>
<td><strong>Assigned Comparison N</strong>: 51</td>
<td><strong>Comparison Type</strong>: Active/Sham</td>
<td><strong>Selective Reporting</strong>: Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Other</strong>: High</td>
</tr>
<tr>
<td>Mabry &amp; Turner (2016)</td>
<td><strong>Educational Setting</strong>: College/University</td>
<td><strong>Design</strong>: RCT</td>
<td><strong>Random Sequence</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>Male</strong>: 100%</td>
<td><strong>Funded</strong>: No</td>
<td><strong>Allocation</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>White</strong>: 65.9%</td>
<td><strong>Peer Reviewed</strong>: Yes</td>
<td><strong>Blinding of Participants</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>Country</strong>: United States</td>
<td><strong>Primary Report Type</strong>: Journal Article</td>
<td><strong>Blinding of Assessment</strong>: High</td>
</tr>
<tr>
<td></td>
<td><strong>Assigned Treatment N</strong>: 109</td>
<td><strong>Treatment Program</strong>: Where Do You Stand</td>
<td><strong>Incomplete Data</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>Assigned Comparison N</strong>: 105</td>
<td><strong>Comparison Type</strong>: Active/Sham</td>
<td><strong>Selective Reporting</strong>: Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Other</strong>: Low</td>
</tr>
<tr>
<td>Miller et al. (2014)</td>
<td><strong>Educational Setting</strong>: Secondary School</td>
<td><strong>Design</strong>: QED</td>
<td><strong>Random Sequence</strong>: High</td>
</tr>
<tr>
<td></td>
<td><strong>Male</strong>: 100%</td>
<td><strong>Funded</strong>: Yes</td>
<td><strong>Allocation</strong>: Unclear</td>
</tr>
<tr>
<td></td>
<td><strong>White</strong>: Not Reported</td>
<td><strong>Peer Reviewed</strong>: Yes</td>
<td><strong>Blinding of Participants</strong>: Unclear</td>
</tr>
<tr>
<td>Study Primary Report</td>
<td>Participants</td>
<td>Methods</td>
<td>Risk of Bias</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------</td>
<td>---------</td>
<td>--------------</td>
</tr>
<tr>
<td><strong>Country:</strong> India</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Assigned Treatment N:</strong> 305</td>
<td><strong>Primary Report Type:</strong> Journal Article</td>
<td><strong>Blinding of Assessment:</strong> High</td>
<td></td>
</tr>
<tr>
<td><strong>Assigned Comparison N:</strong> 276</td>
<td><strong>Treatment Program:</strong> Coaching Boys Into Men</td>
<td><strong>Incomplete Data:</strong> Low</td>
<td></td>
</tr>
<tr>
<td><strong>Method:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Risk of Bias:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Miller et al. (2013)

| **Educational Setting:** Secondary School | **Design:** RCT | **Random Sequence:** Low |
| **Male:** 100% | **Funded:** Yes | **Allocation:** Unclear |
| **White:** 34.0% | **Peer Reviewed:** Yes | **Blinding of Participants:** Unclear |
| **Country:** United States | **Primary Report Type:** Journal Article | **Blinding of Assessment:** High |
| **Assigned Treatment N:** 1,008 | **Treatment Program:** Coaching Boys Into Men | **Incomplete Data:** High |
| **Assigned Comparison N:** 998 | **Comparison Type:** Inactive Comparison | **Selective Reporting:** Unclear |
| **Other:** Low |

Moynihan et al. (2011)

| **Educational Setting:** College/University | **Design:** RCT | **Random Sequence:** Unclear |
| **Male:** 0% | **Funded:** No | **Allocation:** Unclear |
| **White:** Not Reported | **Peer Reviewed:** Yes | **Blinding of Participants:** Unclear |
| **Country:** United States | **Primary Report Type:** Journal Article | **Blinding of Assessment:** High |
| **Assigned Treatment N:** 36 | **Treatment Program:** Bringing in the Bystander | **Incomplete Data:** High |
| **Assigned Comparison N:** 20 | **Comparison Type:** Inactive Comparison | **Selective Reporting:** Low |
| **Other:** High |

Moynihan et al. (2010)

<p>| <strong>Educational Setting:</strong> College/University | <strong>Design:</strong> RCT | <strong>Random Sequence:</strong> Unclear |
| <strong>Male:</strong> 56.7% | <strong>Funded:</strong> No | <strong>Allocation:</strong> Unclear |
| <strong>White:</strong> Not Reported | <strong>Peer Reviewed:</strong> Yes | <strong>Blinding of Participants:</strong> Unclear |
| <strong>Country:</strong> United States | <strong>Primary Report Type:</strong> Journal Article | <strong>Blinding of Assessment:</strong> High |
| <strong>Assigned Treatment N:</strong> 53 | <strong>Treatment Program:</strong> Bringing in the Bystander | <strong>Incomplete Data:</strong> High |
| <strong>Other:</strong> Low | <strong>Comparison Type:</strong> Inactive Comparison | <strong>Selective Reporting:</strong> Low |</p>
<table>
<thead>
<tr>
<th>Study Primary Report</th>
<th>Participants</th>
<th>Methods</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assigned Comparison N:</strong> 86</td>
<td><strong>Comparison Type:</strong> Inactive Comparison</td>
<td><strong>Other:</strong> High</td>
<td></td>
</tr>
<tr>
<td><strong>Educational Setting:</strong> College/University</td>
<td><strong>Design:</strong> QED</td>
<td><strong>Random Sequence:</strong> High</td>
<td></td>
</tr>
<tr>
<td><strong>Male:</strong> 49.1%</td>
<td><strong>Funded:</strong> Yes</td>
<td><strong>Allocation:</strong> Unclear</td>
<td></td>
</tr>
<tr>
<td><strong>White:</strong> 69.9%</td>
<td><strong>Peer Reviewed:</strong> Yes</td>
<td><strong>Blinding of Participants:</strong> Unclear</td>
<td></td>
</tr>
<tr>
<td><strong>Country:</strong> United States</td>
<td><strong>Primary Report Type:</strong> Journal Article</td>
<td><strong>Blinding of Assessment:</strong> High</td>
<td></td>
</tr>
<tr>
<td><strong>Assigned Treatment N:</strong> 369</td>
<td><strong>Treatment Program:</strong> Brining in the Bystander</td>
<td><strong>Incomplete Data:</strong> High</td>
<td></td>
</tr>
<tr>
<td><strong>Assigned Comparison N:</strong> 224</td>
<td><strong>Comparison Type:</strong> Inactive Comparison</td>
<td><strong>Selective Reporting:</strong> Low</td>
<td></td>
</tr>
<tr>
<td><strong>Other:</strong> High</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Educational Setting:** College/University | **Design:** RCT | **Random Sequence:** Unclear |
| **Male:** 48.9% | **Funded:** Yes | **Allocation:** Unclear |
| **White:** Not Reported | **Peer Reviewed:** Yes | **Blinding of Participants:** Unclear |
| **Country:** United States | **Primary Report Type:** Journal Article | **Blinding of Assessment:** High |
| **Assigned Treatment N:** 81 | **Treatment Program:** Know Your Power | **Incomplete Data:** High |
| **Assigned Comparison N:** 64 | **Comparison Type:** Inactive Comparison | **Selective Reporting:** Low |
| **Other:** High |

<p>| <strong>Educational Setting:</strong> College/University | <strong>Design:</strong> RCT | <strong>Random Sequence:</strong> Low |
| <strong>Male:</strong> 100% | <strong>Funded:</strong> Yes | <strong>Allocation:</strong> Low |
| <strong>White:</strong> 44.0% | <strong>Peer Reviewed:</strong> Yes | <strong>Blinding of Participants:</strong> Low |
| <strong>Country:</strong> United States | <strong>Primary Report Type:</strong> Journal Article | <strong>Blinding of Assessment:</strong> High |
| <strong>Assigned Treatment N:</strong> 376 | <strong>Treatment Program:</strong> RealConsent | <strong>Incomplete Data:</strong> High |
| <strong>Assigned Comparison N:</strong> 367 | <strong>Comparison Type:</strong> Active/Sham | <strong>Selective Reporting:</strong> High |
| <strong>Other:</strong> High |</p>
<table>
<thead>
<tr>
<th>Study Primary Report</th>
<th>Participants</th>
<th>Methods</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sargent et al. (2017)</td>
<td><strong>Educational Setting:</strong> Secondary School</td>
<td><strong>Design:</strong> RCT</td>
<td><strong>Random Sequence:</strong> Unclear</td>
</tr>
<tr>
<td></td>
<td>Male: 47.5%</td>
<td><strong>Funded:</strong> Yes</td>
<td><strong>Allocation:</strong> Unclear</td>
</tr>
<tr>
<td></td>
<td>White: 0.3%</td>
<td><strong>Peer Reviewed:</strong> Yes</td>
<td><strong>Blinding of Participants:</strong> Unclear</td>
</tr>
<tr>
<td></td>
<td>Country: United States</td>
<td><strong>Primary Report Type:</strong> Journal Article</td>
<td><strong>Blinding of Assessment:</strong> High</td>
</tr>
<tr>
<td></td>
<td>Assigned Treatment N: 463</td>
<td><strong>Treatment Program:</strong> TakeCARE</td>
<td><strong>Incomplete Data:</strong> Low</td>
</tr>
<tr>
<td></td>
<td>Assigned Comparison N: 458</td>
<td><strong>Comparison Type:</strong> Active/Sham</td>
<td><strong>Selective Reporting:</strong> Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Other:</strong> High</td>
<td><strong>Unclear</strong></td>
</tr>
<tr>
<td>Senn &amp; Forrest (2016)</td>
<td><strong>Educational Setting:</strong> College/University</td>
<td><strong>Design:</strong> QED</td>
<td><strong>Random Sequence:</strong> High</td>
</tr>
<tr>
<td></td>
<td>Male: 25.1%</td>
<td><strong>Funded:</strong> Yes</td>
<td><strong>Allocation:</strong> High</td>
</tr>
<tr>
<td></td>
<td>White: 69.1%</td>
<td><strong>Peer Reviewed:</strong> Yes</td>
<td><strong>Blinding of Participants:</strong> High</td>
</tr>
<tr>
<td></td>
<td>Country: Canada</td>
<td><strong>Primary Report Type:</strong> Journal Article</td>
<td><strong>Blinding of Assessment:</strong> High</td>
</tr>
<tr>
<td></td>
<td>Assigned Treatment N: 545</td>
<td><strong>Treatment Program:</strong> Bringing in the Bystander</td>
<td><strong>Incomplete Data:</strong> High</td>
</tr>
<tr>
<td></td>
<td>Assigned Comparison N: 326</td>
<td><strong>Comparison Type:</strong> Inactive Comparison</td>
<td><strong>Selective Reporting:</strong> Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Other:</strong> Unclear</td>
<td><strong>Unclear</strong></td>
</tr>
</tbody>
</table>
### Table A2: Summary of excluded studies

<table>
<thead>
<tr>
<th>Study Reference(s)</th>
<th>Study Description</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahrens, C. E., Rich, M. D., &amp; Ullman, J. B. (2011). Rehearsing for real life: The impact of the InterACT Sexual Assault Prevention Program on self-reported likelihood of engaging in bystander interventions. <em>Violence Against Women, 17</em>(6), 760-776. doi:10.1177/1077801211410212</td>
<td>Researchers evaluated the InterAct Sexual Assault Prevention program, measuring decisional balance, likelihood of engaging in bystander interventions, in 509 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Alegria-Flores, K., Raker, K., Pleasants, R. K., Weaver, M. A., &amp; Weinberger, M. (2017). Preventing interpersonal violence on college campuses: the effect of one act training on bystander intervention. <em>Journal of interpersonal violence, 32</em>(7), 1103-1126.</td>
<td>The researchers compared the effects of One Act and HAVEN programs, measuring college date rape attitudes and behaviors, bystander confidence, willingness to help, and bystander behavior in 1,487 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment design.</td>
</tr>
<tr>
<td>Amar, A. F., Sutherland, M., &amp; Kesler, E. (2012). Evaluation of a bystander education program. <em>Issues in Mental Health Nursing, 33</em>(12), 851-857. doi:10.3109/01612840.2012.709915</td>
<td>Researchers evaluated a bystander education program, measuring attitudes related to sexual and partner violence and willingness to help in 202 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Baiocchi, M. (2016). A Cluster-randomized Trial to Assess a Sexual Assault Prevention Intervention in Adolescents in Nairobi, Kenya. ClinicalTrials.gov Identifier: NCT02771132.</td>
<td>Researchers evaluated the efficacy of Source of Strength, a bystander intervention, measuring sexual and physical assault incidence.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment study.</td>
</tr>
<tr>
<td>Baldwin-White, A., Thompson, M. S., &amp; Gray, A. (2016). Pre- and Postintervention Factor Analysis of the Illinois Rape Myth Acceptance Scale. <em>Journal of Aggression, Maltreatment &amp; Trauma, 25</em>(6), 636-651. doi:10.1080/10926771.2015.1107173</td>
<td>Researchers evaluated a bystander education intervention, measuring rape myth acceptance in 352 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
</tbody>
</table>

Bannon, R. S. (2014). *The bystander approach to sexual assault risk reduction: effects on risk recognition, perceived self-...*
<table>
<thead>
<tr>
<th>Reference</th>
<th>Title and Description</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcome Measures</th>
</tr>
</thead>
</table>

\textbf{Intervention}: Study does not involve an eligible bystander intervention program. No eligible intervention. |  |  |

\textbf{Comparison}: Study does not have an eligible inactive comparison condition. This is a single-group study. |  |  |

\textbf{Intervention}: Study does not involve an eligible bystander intervention program. Intervention targets school-based peer victimization and bullying, without a sexual assault component. |  |  |

\textbf{Outcome Measures}: Study does not report on at least one outcome related to bystander attitudes/behaviors or IPV/sexual assault perpetration (or outcomes reported by breakout groups that cannot be combined). Study collects qualitative data. |  |  |
| Beardall, N. G. (2008). A program evaluation research study on the implementation of the mentors in violence prevention program in a public high school (Doctoral dissertation, Lesley University). | The researcher evaluated the Mentors in Violence Prevention program, implementing a mixed methods study measuring bystander intentions and attitudes. | 

\textbf{Comparison}: Study does not have an eligible inactive comparison condition. This is a single-group study. |  |  |
<table>
<thead>
<tr>
<th>Study Title</th>
<th>Study Details</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bluth, S. J. (2014). <em>Breaking the culture of silence in the sisterhood: Using bystander intervention</em>. (Doctoral dissertation, Sam Houston State University).</td>
<td>The researcher evaluated the Bringing in the Bystander program, measuring bystander efficacy, rape myths, readiness to help, and attitudes toward gender roles in 13 participants.</td>
<td>Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Bush, H. (2016). <em>ConnectEd: A Randomized Controlled Trial Connecting Through Educational Training (ConnectEd).</em> ClinicalTrials.gov Identifier: NCT02786472.</td>
<td>Researchers evaluated the Green Dot program, measuring bystander behavior, bystander intentions, bystander efficacy, acceptance of violence, and number of risky behaviors.</td>
<td>Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Bush, H. (2015). <em>A Pilot of Bystander Training Programming in Sexual Violence and Substance Abuse (P_ConnectED).</em> ClinicalTrials.gov Identifier: NCT02591212.</td>
<td>Researchers evaluated the Green Dot program, measuring bystander behavior, bystander</td>
<td>Study does not have an eligible inactive comparison condition. This is a treatment-treatment design.</td>
</tr>
<tr>
<td>Reference</td>
<td>Title</td>
<td>Research Design</td>
</tr>
<tr>
<td>Interventions. <strong>Health Promotion Practice</strong>, 10(1), 195. doi:10.1177/1524839908319593</td>
<td>Changes, and intended behaviors in 1,182 participants.</td>
<td>Coker, A. (2016). Multi-College Bystander Efficacy Evaluation. ClinicalTrials.gov Identifier: NCT02659423. Researchers evaluated the Green Dot program, measuring interpersonal victimization, bystander behavior, and violence acceptance. <strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment design.</td>
</tr>
<tr>
<td>Reference</td>
<td>Title</td>
<td>Study Description</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Dempsey, A. G. (2009). <em>Aggression and victimization in middle school: A mixed methods analysis of the process and effectiveness of implementing a prevention program</em> (Order No. 3467697).</td>
<td>The researcher evaluated the Aggressors, Victims, and Bystanders program, measuring victimization, depression, social anxiety, and victim empathy in students attending four schools.</td>
<td><strong>Intervention:</strong> Study does not involve an eligible bystander intervention program. Intervention targets school-based peer victimization and bullying, without a sexual assault component.</td>
</tr>
<tr>
<td>Dick, A. S. (2009). <em>Transforming witnesses to actors: 100+ Men Against Domestic Violence</em>. Rutgers The State University of New Jersey-New Brunswick.</td>
<td>The researcher conducted interviews with seven men involved in 100+ Men Against Domestic Violence, a program intended to increase bystander interventions and alter social norms on campus.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Elias-Lambert, N., &amp; Black, B. M. (2016). <em>Bystander Sexual Violence Prevention Program: Outcomes for High- and Low-Risk University Men</em>. <em>Journal of Interpersonal Violence, 31</em>(19), 3211-3235. doi:10.1177/0886260515584346</td>
<td>Researchers evaluated the efficacy of a bystander intervention program, measuring attitudes and behaviors related to sexual violence in 142 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. Because much of the sample was included in a previous wave of the study, portions of the control group received the intervention.</td>
</tr>
<tr>
<td>El-Khoury, J. R., &amp; Shafer, A. (2016). <em>Narrative Exemplars and the Celebrity Spokesperson in Lebanese Anti-Domestic Violence Public Service Announcements</em>. <em>Journal of Health Communication, 21</em>(8), 935-943.</td>
<td>Researchers evaluated a bystander intervention program that used narrative exemplars and celebrity endorsements, measuring bystander awareness, intervention, and efficacy in 18 classrooms.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment study.</td>
</tr>
<tr>
<td>Espelage, D. (2013). <em>Multi-Site Evaluation of Second Step (SSTP)</em>. ClinicalTrials.gov Identifier: NCT01792167.</td>
<td>Researchers evaluated the effectiveness of Second Step, measuring bullying, sexual harassment perpetration, and dating violence.</td>
<td><strong>Participants and Setting:</strong> Sample not implemented in an educational setting with eligible adolescents or college students. Sample consisted of 6th grade students.</td>
</tr>
<tr>
<td>Foubert, J. D., Godin, E. E., &amp; Tatum, J. L. (2009). In Their Own Words: Sophomore College Men Describe Attitude and Behavior Changes Resulting From a Rape Prevention Program 2 Years After Their Participation. <em>Journal of Interpersonal Violence, 25</em>, 2237-2257. doi:10.1177/0886260509354881</td>
<td>The researchers evaluated The Men’s program, qualitatively exploring attitude and behavior change in 184 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Foubert, J. D., Langhinrichsen-Rohling, J., Brasfield, H., &amp; Hill, B. (2010). Effects of a Rape Awareness Program on College Women: Increasing Bystander Efficacy and Willingness to Intervene. <em>Journal of Community Psychology, 38</em>(7), 813-827.</td>
<td>Researchers evaluated two sexual assault risk-reduction programs, The Men’s Program and the Women’s Program, measuring rape myth acceptance, bystander behavior, and bystander efficacy in 456 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment study, as the members of the control group were required to receive education and training on dating violence and sexual assault.</td>
</tr>
<tr>
<td>Foubert, J. D., &amp; Masin, R. C. (2012). Effects of The Men’s Program on U.S. Army Soldiers’ Intentions to Commit and Willingness to Intervene to Prevent Rape: A Pretest Posttest Study. <em>Violence and Victims, 27</em>(6), 911-921. doi:10.1891/0886-6708.27.6.911</td>
<td>Researchers evaluated the Men’s Project, measuring bystander willingness to help, bystander efficacy, rape myth acceptance, and likelihood of raping in 481 military personnel.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment study.</td>
</tr>
<tr>
<td>Foubert, J. D., &amp; Newberry, J. T. (2006). Effects of Two Versions of an Empathy-Based Rape Prevention Program on Fraternity Men’s Survivor Empathy, Attitudes, and Behavioral Intent to Commit Rape or Sexual Assault. <em>Journal of College Student Development</em>, 133-148. doi:10.1353/csd.2006.0016</td>
<td>Researchers evaluated The Men’s Program, measuring empathy, intent to rape, and rape myth acceptance in 261 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment design.</td>
</tr>
<tr>
<td>Foubert, J. D., &amp; Perry, B. C. (2007). Creating Lasting Attitude and Behavior Change in Fraternity Members and Male Student Athletes: The Qualitative Impact of an Empathy-Based Rape Prevention Program. <em>Violence Against Women, 13</em>, 70-86. doi:10.1177/1077801207002045</td>
<td>The researchers evaluated The Men’s Program, conducting focus groups with 26 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Foubert, J. D., Tatum, J., &amp; Donahue, G. A. (2006). Reactions of First-Year Men to Rape Prevention Program: Attitude and Predicted Behavior Changes. <em>NASPA Journal, 43</em>(3), 578-598.</td>
<td>Researchers evaluated the Men’s Program, measuring bystander attitudes and behaviors in 261 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Reference</td>
<td>Title</td>
<td>Methodology</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>Garman, A. M. (2013).</td>
<td>Increasing the effectiveness of sexual harassment prevention through learner engagement (Doctoral dissertation, California State University, Long Beach).</td>
<td>The researcher evaluated the effectiveness of a bystander intervention, measuring likelihood of sexual harassment in 320 participants.</td>
</tr>
<tr>
<td>Gedney, C. R. (2016).</td>
<td>Sexual assault prevention: A randomized controlled trial of a military intervention (Doctoral dissertation, The University of Utah).</td>
<td>The researcher evaluated the United States Air Force sexual assault prevention program with and without a motivational interview component, measuring bystander attitudes, efficacy, and rape myth acceptance in 51 participants.</td>
</tr>
<tr>
<td>Glass, N. (2014).</td>
<td>Effectiveness of a Safety App to Respond to Dating Violence for College Women and Their Friends. ClinicalTrials.gov Identifier: NCT02236663.</td>
<td>Researchers evaluated the effectiveness of a personalized smart phone and web-based application that creates personalized safety plan for women experiencing intimate partner violence, measuring use of safety strategies, decisional conflict, efficacy to intervene, and supportive behaviors.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Hines, D. A., &amp; Palm Reed, K. M. (2015). An experimental evaluation of peer versus professional educators of a bystander program for the prevention of sexual and dating violence among college students. <em>Journal of Aggression, Maltreatment &amp; Trauma, 24</em>(3), 279-298. doi:10.1080/10926771.2015.1009601</strong></td>
<td>Researchers compared the effectiveness of a peer-led and a professional-led bystander intervention program, measuring sexual violence and domestic violence supporting attitudes, bystander efficacy, and bystander behaviors in 229 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment design.</td>
</tr>
<tr>
<td><strong>Hines, D. A., &amp; Palm Reed, K. M. (2015). Predicting improvement after a bystander program for the prevention of sexual and dating violence. <em>Health Promotion Practice, 16</em>(4), 550-559. doi:10.1177/1524839914557031</strong></td>
<td>Researchers evaluated a bystander prevention program, measuring bystander attitudes, knowledge, and behaviors in 296 participants.</td>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td><strong>Hollingsworth, C., Ramey, K. J., &amp; Hadley, J. A. (2011). Bystander Intervention Pilot: Final Report. Retrieved from <a href="http://www.ncdsv.org/images/USNavy_BystanderInterventionPilot_4-2011.pdf">http://www.ncdsv.org/images/USNavy_BystanderInterventionPilot_4-2011.pdf</a></strong></td>
<td>The researchers evaluated the Mentors in Violence Prevention program, measuring attitudes and beliefs towards sexual assault in 2,918 participants.</td>
<td><strong>Participants and Setting:</strong> Sample not implemented in an educational setting with eligible adolescents or college students. Sample was military, not in an educational setting, and there was no eligible inactive comparison.</td>
</tr>
<tr>
<td>Reference</td>
<td>Title</td>
<td>Study Details</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>---------------</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Title</td>
<td>Summary</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>Kostouros, P., Warthe, G. D., Carter-Snell, C., &amp; Burnett, C. (2016).</td>
<td>&quot;Stepping Up&quot;: A Focus on Facilitator Development. <em>Journal of Student Affairs Research and Practice</em>, 53(2), 218-229.</td>
<td>Researchers explored facilitators’ experiences implementing a bystander program, conducting a focus group with eight participants.</td>
</tr>
<tr>
<td>Liu, E. H. (2010).</td>
<td>A study of a university-based men-only prevention program (men care): Effect on attitudes and behaviors related to sexual violence. (Doctoral dissertation, University of Southern California.)</td>
<td>The researcher evaluated Men CARE, measuring attitudes and behaviors toward sexual violence in 901 participants.</td>
</tr>
<tr>
<td>Reference</td>
<td>Methodology</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>McMahon, S., Winter, S. C., Palmer, J. E., Postmus, J. L., Peterson, A. N., Zucker, S., &amp; Koenick, R. (2015). A Randomized Controlled Trial of a Multi-Dose Bystander Intervention Program Using Peer Education Theater. <em>Health Education Research</em>, 30(4), 554-568.</td>
<td>The researchers evaluated Students Challenging Realities and Education Against Myths (SCREAM) program, measuring bystander outcomes in 1,390 participants, a random subset of whom received three treatment doses, and the control group received only one.</td>
<td>Comparison: Study does not have an eligible inactive comparison condition. Both treatment and comparison group received treatment; treatment group received additional doses.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Research Focus</th>
<th>Participants and Setting</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nabity, J. (2010).</td>
<td>The bystander approach to sexual harassment training: Considering a new perspective (Order No. 1490318). Available from ProQuest Dissertations &amp; Theses Global. (860136402).</td>
<td>The researcher compared two sexual harassment training videos, one using bystander intervention language and one using accusatory language, to determine differences in likelihood of future sexual harassment and acceptance of sexual harassment myths in 189 participants.</td>
<td>Intervention: Study does not involve an eligible bystander intervention program. Study assesses a bystander program to reduce sexual harassment in the workplace, not sexual assault.</td>
<td></td>
</tr>
<tr>
<td>Nitzel, C. L. (2016).</td>
<td>The process of engagement with the green dot bystander intervention program among higher education staff/faculty (Order No. 10102750).</td>
<td>The researcher explored faculty and staff views of the Green Dot bystander program.</td>
<td>Comparison: Study does not have an eligible inactive comparison condition. This is a single-group qualitative study.</td>
<td></td>
</tr>
<tr>
<td>Study Title</td>
<td>Authors</td>
<td>Intervention Design</td>
<td>Comparison</td>
<td>Key Findings</td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>---------------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Potter, S. J. (2012). Using a Multimedia Social Marketing Campaign to Increase Active Bystanders on the College Campus. <em>Journal of American College Health, 60</em>(4), 282-295.</td>
<td>The researcher evaluated the Know Your Power program, measuring community perceptions of relationship problems in 353 participants.</td>
<td>Comparison: Study does not have an eligible inactive comparison condition. No inactive comparison group, as all participants were exposed to the treatment and were then divided into groups based on whether they recalled seeing the treatment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potter, S. J., Moynihan, M. M. (2011). Bringing in the bystander in-person prevention program to a U.S. military installation: Results from a pilot study. <em>Military Medicine, 176</em>(8), 870-875. doi:10.7205/MILMED-D-10-00483</td>
<td>Researchers evaluated the Bringing in the Bystander program on a military base, measuring bystander behavior in 394 participants.</td>
<td>Participants and Setting: Sample not implemented in an educational setting with eligible adolescents or college students. Participants are military personnel.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potter, S. J., Moynihan, M. M., &amp; Stapleton, J. G. (2011). Using Social Self-Identification in Social Marketing Materials Aimed at Reducing Violence against Women on Campus. <em>Journal of Interpersonal Violence, 26</em>(5), 971-990.</td>
<td>Researchers evaluated a poster campaign, measuring willingness to engage in bystander behavior in 372 participants.</td>
<td>Comparison: Study does not have an eligible inactive comparison condition. No inactive comparison group, as all participants were exposed to the treatment and were then divided into groups based on whether they recalled seeing the treatment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potter, S. J., Moynihan, M. M., Stapleton, J. G., &amp; Banyard, V. L. (2009).</td>
<td>Researchers evaluated a poster campaign, measuring prosocial bystander behaviors and willingness to intervene in 372 participants.</td>
<td>Comparison: Study does not have an eligible inactive comparison condition. No inactive comparison group, as all participants were exposed to the treatment and were then divided into groups based on whether they recalled seeing the treatment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potter, S. J., &amp; Stapleton, J. G. (2013). Assessing the efficacy of a bystander social marketing campaign four weeks following the campaign administration. <em>Sexual Assault Report, 16</em>, 65-80.</td>
<td>Researchers evaluated the Know Your Power program, measuring rape myth acceptance and readiness to change in 236 participants.</td>
<td>Comparison: Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Potter, S. J., &amp; Stapleton, J. G. (2012). Translating Sexual Assault Prevention from a College Campus to a United States Military Installation: Piloting the Know-Your-Power Bystander Social Marketing Campaign. <em>Journal of Interpersonal Violence, 27</em>(8), 1593-1621.</td>
<td>Researchers evaluated the Bringing in the Bystander and Know Your Power programs, measuring bystander behavior, bystander confidence, and awareness of role in preventing sexual assault in 155 participants.</td>
<td>Participants and Setting: Sample not implemented in an educational setting with eligible adolescents or college students. Participants are members of the US military.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schumitsch-Jewell, T (2016). Stand Up and Speak Up: Effectiveness of Web-Based Bystander Intervention on a College Campus. <em>Oshkosh Scholar, 11</em>, 8-20.</td>
<td>The researcher evaluated a bystander intervention, measuring rape myth acceptance, bystander efficacy, and bystander confidence in 28 participants.</td>
<td>Comparison: Study does not have an eligible inactive comparison condition. This is a treatment-treatment study.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Title</td>
<td>Researcher(s)</td>
<td>Methodology</td>
<td>Comparison</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>---------------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>Shifflet, J. H. (2013).</td>
<td>Action research evaluation of bystander intervention training created by Munche, Stern, and O’Brien (Doctoral dissertation, Capella University).</td>
<td>The researcher conducted a qualitative investigation of a bystander intervention training with 15 military personnel.</td>
<td></td>
<td>Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Sparks, K. B. (2015).</td>
<td>Initiating a Bystander Awareness Program at a State University (Doctoral dissertation, Seton Hall University).</td>
<td>The researcher evaluated the Step UP program, measuring violence witnessed and opinions on intervention in 150 participants.</td>
<td></td>
<td>Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Stephens, M. R. (2013).</td>
<td>Helping survivors of sexual assault: The role of general and event-specific empathy (Doctoral dissertation, California State University, Long Beach).</td>
<td>The researcher evaluated the InterAct program, measuring empathy and helping behavior in 393 participants.</td>
<td></td>
<td>Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
</tr>
<tr>
<td>Taylor, B. et al. (2011).</td>
<td>Shifting boundaries: final report on an experimental evaluation of a youth dating violence prevention program in New York City middle schools. National Institute of Justice; Washington, DC.</td>
<td>Researchers evaluated the Shifting Boundaries program, measuring sexual harassment perpetration, sexual dating violence, behavioral intentions, and dating violence knowledge in 117 classrooms.</td>
<td></td>
<td>Sample not implemented in an educational setting with eligible adolescents or college students. Mean age of participants is less than 12.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The researchers evaluated the Shifting Boundaries program, measuring sexual harassment victimization and perpetration in 1,764 students.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a treatment-treatment study.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Researchers evaluated the Shifting Boundaries program, measuring sexual and non-sexual violence harassment victimization and perpetration in 123 classrooms.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Participants and Setting:</strong> Sample not implemented in an educational setting with eligible adolescents or college students. Mean age of participants was less than 12.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Researchers evaluated the Shifting Boundaries program, measuring sexual and non-sexual violence harassment victimization and perpetration in 123 classrooms.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Participants and Setting:</strong> Sample not implemented in an educational setting with eligible adolescents or college students. Mean age of participants was less than 12.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Researchers explore bystander roles, outlining a theoretical framework and exploring teachers’ perceptions of bullying.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intervention:</strong> Study does not involve an eligible bystander intervention program. This is a non-empirical piece.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevention (pp. 215-232): New York Academy of Sciences, New York, NY.</td>
<td>Vadovic, R. J. (2013). <em>Implementing a bystander awareness program on a university campus</em> (Order No. 3590174). The researcher evaluated the Step Up program, measuring exposure to violence, victimization, response to violence, and social norms around violence in 236 participants. <strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Warthe, D. G., &amp; Kostouros, P. (2015). Stepping Up: Reducing Dating Violence on Post-Secondary Campuses. Annual Conference - End Violence Against Women International Researchers evaluated the Stepping Up program, measuring knowledge in 37 participants. <strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weitzman, A., Cowan, S., &amp; Walsh, K. (2017). Bystander Interventions on Behalf of Sexual Assault and Intimate Partner Violence Victims. <em>Journal of Interpersonal Violence, 0886260517696873.</em> Researchers report the results of a longitudinal study, measuring knowledge, intervention, and barriers to intervention in 1,307 participants. <strong>Intervention:</strong> Study does not involve an eligible bystander intervention program. No intervention evaluated.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams, D. J., &amp; Neville, F. G. (2017). Qualitative evaluation of the mentors in violence prevention pilot in Scottish high schools. <em>Psychology of violence, 7</em>(2), 213. Researchers evaluated the Mentors in Violence Prevention program, conducting interviews and focus groups to explore experiences with the program. <strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winegust, A. K. (2015). <em>Pass it On: An Evaluation of a Sexualized Violence Prevention Program for Middle School and High School Students</em> (Doctoral dissertation, University of Toronto (Canada)). The researcher evaluated the Pass It On program, a feminist bystander intervention program, <strong>Comparison:</strong> Study does not have an eligible inactive comparison condition. This is a single-group study.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
measuring both bystander and victim blaming attitudes in 152 participants.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study Title</th>
<th>Authors</th>
<th>Methodology</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woglom, L., &amp; Pennington, K. (2010). The Bystander’s Dilemma: How Can We Turn Our Students into Upstanders? Social Education, 74(5), 254-258.</td>
<td>Researchers draw from various theoretical frameworks to discuss ways of cultivating bystander intervention behavior in their students.</td>
<td><strong>Intervention</strong>: Study does not involve an eligible bystander intervention program. Researchers did not conduct an intervention; this is a theoretical piece.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yoshihama, M., &amp; Tolman, R. M. (2015). Using interactive theater to create socioculturally relevant community-based intimate partner violence prevention. American Journal of Community Psychology, 55(1-2), 136-147. doi:10.1007/s10464-014-9700-0</td>
<td>Researchers describe the implementation of a theater program, discussing the benefits, challenges, and limitations of the program.</td>
<td><strong>Intervention</strong>: Study does not involve an eligible bystander intervention program. Researchers describe implementation but do not assess the intervention.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Appendix: Codebook

## BYSTANDER INTERVENTION
### META-ANALYSIS CODING MANUAL

Last Updated May 17, 2017

### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligibility Criteria</td>
<td>133</td>
</tr>
<tr>
<td>Full-Text Coding</td>
<td>136</td>
</tr>
<tr>
<td>Study Level</td>
<td>136</td>
</tr>
<tr>
<td>Intervention and Comparison Groups</td>
<td>146</td>
</tr>
<tr>
<td>Outcomes</td>
<td>152</td>
</tr>
<tr>
<td>Effect Sizes</td>
<td>154</td>
</tr>
</tbody>
</table>
Eligibility criteria

**Intervention**

1. Eligible intervention programs are those that approach participants as allies in preventing and/or alleviating intimate partner violence (IPV) and/or sexual assault among adolescents and/or college students. (NOTE: For the purposes of this meta-analysis, IPV includes physical or sexual violence perpetrated by an intimate partner. Sexual assault refers to sexual violence that may be perpetrated by an intimate partner, acquaintance, or stranger). Some part of the program must focus on ways that cultivate a willingness for a person to respond to others who are at risk for IPV or sexual assault. All delivery formats are eligible for inclusion (e.g., in-person training sessions, video programs, web-based training, advertising/poster campaigns). There are no treatment duration criteria for inclusion. The following interventions are ineligible:

   a. Studies that report bystander outcomes but do not meet the aforementioned intervention criteria are not eligible for inclusion. For example, studies that assess bystander outcomes of general IPV/sexual assault prevention programs that contain NO bystander content.

   b. Studies that assess outcomes of programs that aim to facilitate pro-social bystander behavior, but that do not explicitly include a component addressing IPV and/or sexual assault (e.g., programs that use a bystander approach to prevent bullying) are not eligible.

**COMPARISON**

2. Eligible comparison groups must receive no intervention services targeting bystander attitudes/behavior, IPV, or sexual assault. Thus, treatment-treatment studies that compare individuals assigned to a bystander program with individuals assigned to a general IPV or sexual assault prevention program are not eligible for inclusion. However, eligible comparison groups may receive a sham or attention treatment that is expected to have no effect on bystander outcomes or attitudes/behaviors regarding IPV or sexual assault (e.g., nutrition or health education).

**PARTICIPANTS AND SETTING**

3. The intervention must be implemented with adolescents and/or college students in educational settings. Eligible participants include adolescents enrolled in grades 7 through 12 and college students enrolled in any type of undergraduate postsecondary educational institution. Studies
that assess bystander programs implemented with adolescents and young adults outside of educational institutions (e.g., in community settings, military settings) are ineligible for inclusion in the review.

a. Eligible studies include those that report on general samples of adolescents and/or college students as well as studies using specialized samples such as those primarily consisting of college athletes, fraternity/sorority members, and single-sex samples.
b. Study samples primarily consisting of post-graduate students will be ineligible for inclusion; the mean age of samples may be no greater than 25 to be included in the review.

**RESEARCH DESIGN**

4. Eligible studies must use an experimental or controlled quasi-experimental research design to compare an intervention group (e.g., students assigned to complete a bystander program) with a comparison group (e.g., students not assigned to complete a bystander program). The following are eligible designs:

a. Randomized controlled trials: Studies in which individuals, classrooms, schools, or other groups are randomly assigned to intervention and comparison conditions.
b. Quasi-randomized controlled trials: Studies where assignment to conditions is quasi-random, for example, by birth date, day of week, student identification number, month, or some other alternation method.
c. Controlled quasi-experimental designs: Studies where participants are not assigned to conditions randomly or quasi-randomly (e.g., participants self-select into groups). Given the potential selection biases inherent in these controlled quasi-experimental designs, we will only include those that also meet one of the following criteria:
   i. Regression discontinuity designs: Studies that use a cutoff on a forcing variable to assign participants to intervention and comparison groups, and assess program impacts around the cutoff of the forcing variable.
   ii. Studies that use propensity score or other matching procedures to create a matched sample of participants in the intervention and comparison groups. To be eligible for inclusion, these studies must also provide enough statistical information to permit estimation of pretest effect sizes for the matched groups (NOTE: Pretest effect sizes refer to pre-intervention measures of an outcome variable. Pre-intervention measures that are not reported as outcomes are group equivalence measures).
   iii. For studies where participants in the intervention and comparison groups are not matched, enough statistical information must be reported that will permit us to estimate pretest effect sizes for at least one outcome measure. Pretest must be administered to the treatment and comparison group at the same point in time.

**SAMPLE SIZES**

5. Intervention and control groups must each contain at least 10 individuals at the time of assignment to study conditions.
OUTCOME MEASURES

6. Eligible studies must report the effects of bystander programs on at least one of the following primary outcomes:
   
a. General attitudes toward IPV and/or sexual assault and victims (e.g., victim empathy, rape myth acceptance).
b. Self-efficacy with regard to bystander intervention (e.g., measures of confidence in respondents’ ability to intervene).
c. Intentions to intervene when witnessing instances or warning signs of IPV and/or sexual assault.
d. Actual intervention behavior when witnessing instances or warning signs of IPV and/or sexual assault.
e. Perpetration of IPV and/or sexual assault.

NOTES: (1) Studies that report outcomes by breakout groups (e.g., report findings separately for males/females, fraternity/nonfraternity, etc.) are not eligible unless findings are reported such that outcomes for breakout groups can be combined. (2) Studies must assess at least one outcome pertinent to sexual assault to be included in the main Campbell review and meta-analysis. Studies that only assess physical violence outcomes (e.g., evaluations of IPV bystander programs that measure physical violence outcomes, but not sexual violence outcomes) will be coded and set aside for separate analysis. These must be marked as “Physical Violence Only” on the eligibility screen in FileMaker. There are no other restrictions on eligibility. Studies that meet all eligibility criteria but do not provide sufficient information for calculating effect sizes are still eligible for inclusion and will be coded (i.e., the study will be coded on all items except the effect sizes).
Full-text coding

Study level

[studyid]
Study identification number. The “unit” you will code here consists of a study, i.e., one research investigation of a defined subject sample or subsamples compared to each other, and the treatments, measures, and statistical analyses applied to them. Sometimes there are several different reports about a single study. In such cases, the coding should be done from the full set of relevant reports, using whichever report is best for each item to be coded; BE SURE YOU HAVE THE FULL SET OF RELEVANT REPORTS BEFORE BEGINNING TO CODE. Sometimes a single report describes more than one study sample, e.g., evaluations at three separate sites. In these cases, each study sample will have a unique study identification number and each study should be coded separately as if it had been described in a separate report.

Each study has its own study identification number, or StudyID (e.g., 619). Each report also has an identification number (e.g., 619.01), which you will find in the FileMaker bibliography. The ReportID has two parts; the part before the decimal is the StudyID, and the part after the decimal is used to distinguish the reports within a study. (These two types of ID numbers, along with bibliographic information, are assigned and tracked using the bibliography.) When coding, use the study ID (e.g., 619) to refer to the study as a whole, and use the appropriate report ID (e.g., 619.01) when referring to an individual report.

[coder]
Coder's initials

HEADER VARIABLES

[pub.year]
Year of publication for the primary report (four digits). If more than one report exists, choose the date for the report that provides the effect sizes. If effect sizes come from more than one report, choose the earliest date.

[publication]
Type of publication for the primary report. If you are using more than one type of publication to code your study, choose the publication that supplies the effect sizes (in cases where more than one
report provides effect sizes, choose a “peer reviewed” choice over another option, or choose the report that provides the most effect sizes).

1. Journal Article
2. Book or book chapter
3. Dissertation
4. MA/MS Thesis
5. Private report
6. Government report
7. Conference paper
8. Other
9. Unclear

[peer]
Is this a peer-reviewed publication (primary report only)?
1. Yes
2. No
3. Unclear

[funded]
Do the authors declare any funding?
1. Yes
2. No
3. Unclear

[researcher_role]
Did one or more of the researchers assessing outcomes of the intervention program (i.e., authors of the research report) play a significant role in developing the program?
1. Yes – One or more researchers developed the intervention program.
2. No – Researchers are independent evaluators.
3. Unclear/Cannot tell.

The following are the developers of some potentially eligible bystander (or peer/social network-based) programs. Some of the programs below were developed by individuals (e.g., academics) and others were developed by organizations. In the latter case, any research report authored by a representative of the developer/organization should be coded as “1.” Less popular programs that do not appear on this list may require a careful reading of the text (e.g., the authors will likely cite the program in the evaluation report) and/or some additional research.

- **Bringing in the Bystander**: Developed by Victoria Banyard, Mary Moynihan, Elizabeth Plante (University of New Hampshire).
- **Circle of 6 (potentially ineligible)**: Developed by Nancy Schwartzman, Thomas Cabus. See list of advisory board members (http://www.circleof6app.com/about/#member2).
- **Coaching Boys into Men**: Developed by Futures without Violence. See list of board and staff members (https://www.futureswithoutviolence.org/about-us/board-and-staff/).
- **Green Dot**: Developed by Dorthy Edwards.
- **Hollaback (potentially ineligible)**: Developed by Emily May.
- **InterACT**: Marc Rich may or may not be the developer, but he has a conflict of interest, as he directs this program/troupe.
- **Know Your Power**: This is the campaign companion to Bringing in the Bystander. Developed by Sharyn Potter, Jane Stapleton, Mary Moynihan (University of New Hampshire).
- **Men Can Stop Rape**: Developed by John Stoltenberg. See MCSR staff members (http://www.mencanstoprape.org/Our-Staff/) and board members (http://www.mencanstoprape.org/Our-Board/).
- **Mentors in Violence Prevention (MVP)**: Developed by Jackson Katz.
- **Speak Up Speak Out (potentially ineligible)**: Developed by Maryland Coalition Against Sexual Assault.
- **Step Up!**: Developed by Becky Bell.
- **That’s Not Cool**: Developed by Futures without Violence. See list of board and staff members (https://www.futureswithoutviolence.org/about-us/board-and-staff/).
- **The Men’s Program (How to Help a Sexual Assault Survivor)**: Developed by John Foubert (Oklahoma State University).
- **The Women’s Program**: Developed by John Foubert (Oklahoma State University).
- **Where Do You Stand**: This is an ad campaign developed by Men Can Stop Rape. MCSR was developed by John Stoltenberg. See MCSR staff members (http://www.mencanstoprape.org/Our-Staff/) and board members (http://www.mencanstoprape.org/Our-Board/).
- **White Ribbon Campaign**: Developed by Michael Kaufman, Jack Layton.

**GROUP IDENTIFICATION AND SELECTION**
At this stage, you will need to identify the groups in the study for which effect size statistics can be computed. Note that for any group comparison coding, the two groups involved must be from the same experiment or quasi-experiment; that is, they must have been involved in the same randomization, matching, etc. from the same design. If two or more experiments or quasi-experiments are presented in the same report, each must be handled separately.

**[txa-d]**
Treatment Groups 1-4 ______________________________________________________________________

**[cta-d]**
Comparison Groups 1-4 ______________________________________________________________________

**STUDY CONTEXT**
[country]
Country in which the study was conducted:
   1. United States
   2. Canada
   3. Mexico
   4. Australia
   5. European Nation
   6. Other

[state]
State in which the study was conducted.
   1. Alabama
   2. ...
   3. ...
   51. District of Columbia
   52. Single state (unspecified)
   53. Multiple states (unspecified)
   54. NA – Conducted outside US

[study.year]
Year(s) when study was conducted. Code NA if cannot tell.

[ed.setting]
Type of educational setting in which the study was conducted:
   1. Secondary school (grades 7-12)
   2. College or university
   3. Unclear

STUDY DESIGN AND METHODS
[design]
Method of assignment to groups. This item focuses on the initial method of assignment to groups regardless of subsequent degradations due to attrition, refusal, etc. prior to treatment onset. These latter situations are coded elsewhere.

Random or near-random:
   1. Random assignment at the individual level: Individual participants are randomly assigned to conditions. In some cases random assignment may be done after individuals have been matched or blocked.
   2. Random assignment by group: Intact groups such as classrooms are assigned.
   3. Regression discontinuity design: Individuals are assigned to groups based on a cut-off score of a forcing variable (i.e., pretest or risk factor measure). Note that this design will be rare.
   4. Quasi-randomized procedure presumed to produce comparable groups: This applies to groups which have individuals assigned by some naturally occurring process that is
apparently random, e.g. alternation, date of birth, medical record number. The key here is that the procedure used to select groups is not strictly random, but the method of allocation should not create nonequivalence between groups.

**Non-random, but matched or statistically controlled:**
Note: Matching refers to the process by which individuals are selected for conditions (e.g., treatment and comparison) in a manner that ensures that the individuals in one group are matched on certain relevant characteristics in the other group. Comparing the characteristics of the groups AFTER they have been assigned to experimental conditions does NOT constitute matching.

5. Matched individually, through sampling, on one or more pretest measures (i.e., attitudes toward IPV/sexual assault or victims, bystander self-efficacy, bystander intentions, bystander behavior, IPV/sexual assault perpetration).

6. Participants in the intervention and comparison groups are not matched, but enough statistical information is provided to permit the estimation of pretest effect sizes for at least one outcome measure (i.e., attitudes toward IPV/sexual assault or victims, bystander self-efficacy, bystander intentions, bystander behavior, IPV/sexual assault perpetration).

7. Matched at a larger group level; that is, intact groups were matched on their means for one or more pretest measures (i.e., attitudes toward IPV/sexual assault or victims, bystander self-efficacy, bystander intentions, bystander behavior, IPV/sexual assault perpetration). In these cases the mean pretest measures of the groups are similar, but each subject in one group has not been individually matched to a subject in the other group on age.

[matched.var.list]
Please list all of the variables used in the matching and/or statistical controls.

[m]
For cluster randomized trials, please enter the average cluster size (i.e., average number of participants in each cluster). Code -9 for cannot tell. Code -8 for Not applicable.

[attrf_o]
What is the overall attrition rate (across all groups) in the study between the time of assignment to groups to the first follow-up? This item refers to overall attrition in the study; more detailed attrition calculations will be estimated using the assigned and observed sample sizes coded in the effect size section.

[attrl_o]
What is the overall attrition rate (across all groups) in the study between the time of assignment to groups to the last follow-up? Again, this item refers to overall attrition in the study; more detailed attrition calculations will be estimated using the assigned and observed sample sizes coded in the effect size section.

[itt]
Did the authors use an intent-to-treat (ITT) analysis? Intent-to-treat analysis refers to situations where researchers ‘analyze as randomized’, meaning that all individuals who were randomized to
the intervention/control groups are included in the final outcome analysis, regardless of whether they actually attended/completed the intervention. Note, that it is possible for a study to conduct an ITT analysis even if they have attrition, as long as they had intended to include any non-completers in their final model.

1. Yes – explicit ITT analysis
2. No – completer analysis, TOT analysis
3. Cannot tell

[missdata]
How did the authors handle missing data in their analysis?

1. Listwise deletion
2. Pairwise deletion
3. Mean or mode imputation
4. Single regression imputation
5. Dummy variable approach (imputed value at zero with dummy variable)
6. Multiple imputation
7. Full information maximum likelihood (FIML)
8. Other method
9. Not applicable – no missing data
10. Cannot tell

**RISK OF BIAS: NON-EXPERIMENTAL STUDIES**
Although these risk of bias variables will largely be used in analyses of non-experimental studies, please code them for all eligible studies, regardless of research design.

[rob_ne_confounding]
Describe any risk of bias for non-experimental studies based on pre-specified confounding factors (i.e., gender, fraternity/sorority membership, athletic team membership, pre-intervention attitudes, pre-intervention bystander measures, and prior sexual assault victimization). That is, please note if a considerable percentage of participants exhibited any of these confounding variables (e.g., if the study was implemented with a sample of fraternity/sorority members).

[rob_ne_cointerventions]
Describe any risk of bias for non-experimental studies based on pre-specified co-interventions (i.e, general sexual assault prevention programs, IPV prevention programs, and general bystander programs). That is, please note if there is reason to believe participants had previous or concurrent exposure to any potentially confounding co-interventions.

**RISK OF BIAS: EXPERIMENTAL STUDIES**
Although these risk of bias variables will largely be used in analyses of experimental studies, please code them for all eligible studies, regardless of research design.

[rob_sg]
Random Sequence Generation: What is the risk of sequence generation due to inadequate generation of a randomized sequence?

1. Low risk. The investigators describe a random component in the sequence generation process such as:
   a. Referring to a random number table;
   b. Using a computer random number generator;
   c. Coin tossing;
   d. Shuffling cards or envelopes;
   e. Throwing dice;
   f. Drawing of lots;
   g. Minimization (may be implemented without a random element, and this is considered equivalent to being random).

2. High risk. The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic non-random approach, for example:
   a. Sequence generated by odd/even birth dates;
   b. Sequence generated by some rule based on date (or day) of admission;
   c. Sequence generated by some rule based on some sort of (e.g., hospital/clinic) record number.

Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgment or some method of non-random categorization of participants, for example:
   a. Allocation by judgment of the clinician (or teacher or practitioner);
   b. Allocation by preference of the participant;
   c. Allocation based on results of a laboratory test or series of tests;
   d. Allocation by availability of the intervention.

By definition, any quasi-experimental design where participants self-select into conditions is at high risk of bias.

3. Unclear risk of bias. Insufficient information is provided about the sequence generation process to permit judgement of low or high risk.

[rob_sg_text]
Provide a description of the information used to code the risk of bias due to sequence generation.

[rob_allocation]
Allocation Concealment: What is the risk of bias due to allocation concealment (inadequate concealment of allocations prior to assignment)? Note that most studies will not report information on allocation concealment.

1. Low risk. Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:
   a. Central allocation (including telephone, web-based, and pharmacy-controlled randomization);
   b. Sequentially numbered drug containers of identical appearance;
   c. Sequentially numbered, opaque, sealed envelopes
2. High risk. Participants or investigators enrolling participants could possibly foresee assignments and thus introduce bias, such as allocation based on:
   a. Using an open random allocation schedule (e.g., a list of random numbers);
   b. Assignment envelopes were used without appropriate safeguards (i.e., if envelopes were unsealed or non-opaque or not sequentially numbered);
   c. Alternation or rotation;
   d. Date of birth;
   e. Case record number;
   f. Any other explicitly unconcealed procedure.

3. Unclear risk. Insufficient information to permit judgment of ‘low risk’ or ‘high risk.’ This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque or sealed.

[rob_allocation_text]
Provide a description of the information used to code the risk of bias due to allocation concealment.

[rob_blinding_part]
Blinding of Participants and Personnel: What is the risk of bias due to knowledge of the allocated interventions by participants and personnel?
1. Low risk. Any one of the following:
   a. No blinding or incomplete blinding, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding;
   b. Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken.
2. High risk. Any one of the following:
   a. No blinding or incomplete blinding, and the outcome measurement is likely to be influenced by lack of blinding;
   b. Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding.
3. Unclear risk. Insufficient information to permit judgment of ‘low risk’ or ‘high risk.’

[rob_blinding_part_text]
Provide a description of the information used to code the risk of bias due to insufficient blinding of participants and personnel.

[rob_blinding_assess]
Blinding of Outcome Assessment: What is the risk of bias due to knowledge of the allocated interventions by outcome assessors?
1. Low risk. Any one of the following:
   a. No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding;
b. Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken

2. High risk. Any one of the following:
   a. No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding;
   b. Blinding of outcome assessment, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding.
   c. All outcomes are self-reported outcomes (which inherently are NOT blinded)

3. Unclear risk. Insufficient information to permit judgment of ‘low risk’ or ‘high risk.’

[rob_blinding_assess_text]
Provide a description of the information used to code the risk of bias due to insufficient blinding of outcome assessment.

[rob_incomplete]
Incomplete Outcome Data: What is the risk of attrition bias due to amount, nature, or handling of incomplete outcome data?

1. Low risk. Any one of the following:
   a. No missing outcome data;
   b. Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias);
   c. Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups;
   d. For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate;
   e. For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size;
   f. Missing data have been imputed using appropriate methods.

2. High risk. Any one of the following:
   a. Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups;
   b. For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate;
   c. For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size;
   d. “As-treated” analysis done with substantial departure of the intervention received from that assigned at randomization;
   e. Potentially inappropriate application of single imputation.

3. Unclear risk. Insufficient information to permit judgment of ‘low risk’ or ‘high risk.’

[rob_incomplete_text]
Provide a description of the information used to code the risk of bias due to incomplete outcome data.

[rob_report]

**Selective Reporting:** What is the risk of bias due to *selective outcome reporting*?

1. **Low Risk of Bias.** Any one of the following:
   a. The study protocol is available and all of the study’s pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way;
   b. The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon)

2. **High Risk of Bias.** Any one of the following:
   a. Not all of the study’s pre-specified primary outcomes have been reported;
   b. One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified;
   c. One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect);
   d. One or more outcomes of interest in the review are reported incompletely so that they cannot be entered into a meta-analysis;
   e. The study report fails to include results for a key outcome that would be expected to have been reported for such a study.

3. **Unclear/Cannot Tell.** Insufficient information to permit judgment of ‘low risk’ or ‘high risk.’

[rob_report_text]

Provide a description of the information used to code the risk of bias due to selective reporting.

[rob_other]

**Other Bias:** What is the risk of bias due to *problems not covered elsewhere*?

1. **Low risk.** The study appears to be free of other sources of bias.

2. **High risk.** There is at least one important risk of bias. For example, the study:
   a. Had a potential source of bias related to the specific study designed used; or
   b. Has been claimed to be fraudulent; or
   c. Had some other problem.

3. **Unclear/cannot tell.** Insufficient information to assess whether an important risk of bias exists.

[rob_other_text]

Provide a description of the information used to code the risk of bias due to problems not covered elsewhere.
Intervention and comparison groups

Create one record in this database for each of the intervention and/or comparison groups you selected earlier for coding. For example, studies with a single intervention group and a single comparison group will have two records in this section of the database.

GROUP IDENTIFICATION
[groupid]
Number each group consecutively within a study, starting with 1.

[tvc]
Select the type of group you are coding.
1. Intervention group
2. Control/comparison group

[type]
What type of services does this group receive?
1. Focal/primary bystander intervention program. There may be several focal programs in a study, as when two different types of programs are compared, both of which are expected to be effective.
2. Active treatment that is not a bystander program or IPV/sexual violence prevention program. This is a group that receives a sham treatment (e.g., watches a video on nature, receives nutrition information, diet intervention) intended to take the same duration as the focal intervention program, but does not involve any active bystander or IPV/sexual violence prevention/education components.
3. Inactive comparison condition. This is a group that receives no prevention program and gets only assessments.
4. Active business as usual. This is a group that receives “usual” active treatment (e.g., school-based sex education, college orientation that contains some sex education) that may be effective in preventing sexual violence but is not the focal treatment of the study. This treatment must be limited to services that the group would receive whether or not the research study was implemented (e.g., mandated college orientation). Programs that focus specifically on bystander behaviors or IPV/sexual assault are not eligible controls; however, sex education programs that may contain limited information on violence are eligible.
5. Other.

[name]
Program name. Write in the program name or label for this group. Please use the exact name the study authors report (many of the programs will be branded – Bringing in the Bystander, Mentors in Violence Prevention).

[descrip]
Program description. Write in a brief description of the treatment this group receives. As much as possible, quote or give a close paraphrase of the relevant descriptive text in the study report; always include page numbers from the report when appropriate. It is acceptable to copy and paste directly from the article as long as you include the information in quotations and provide a page number for the quotation.
Participant Characteristics

[permale]
Enter the percent of participants who identify as male in this group. Use -9 for cannot tell.

[perfemale]
Enter the percent of participants who identify as female in this group. Use -9 for cannot tell.

[pertrans]
Enter the percent of participants who identify as transgender in this group. Use -9 for cannot tell.

[perwhite]
Enter the percent of White participants in this group. Use -9 for cannot tell.

[pernonwhite]
Enter the percent of Non-White participants in this group. Use -9 for cannot tell.

[perblack]
Enter the percent of Black participants in this group. Use -9 for cannot tell.

[perhisp]
Enter the percent of Hispanic participants in this group. Use -9 for cannot tell.

[perasian]
Enter the percent of Asian participants in this group. Use -9 for cannot tell.

[age]
Enter the average age of the group using number of years. Use -9 for cannot tell.

[agerange]
Enter the age range of the group using “XX-XX” format. Use -9 for cannot tell.

[ed_level]
What is the grade/educational level for the majority of the group?

1. 7th-8th grade
2. 9th-10th grade
3. 11th-12th grade
4. First year college
5. Second year college
6. Third year college
7. Fourth year college
8. Other/Cannot tell

[perfrat]
Enter the percent of participants in this group who reported being a member of a fraternity. Use -9 for cannot tell.

[persor]
Enter the percent of participants in this group who reported being a member of a sorority. Use -9 for cannot tell.

[pergreek]
Enter the percent of participants in this group who reported being a member of a fraternity or sorority. Note that this item is of value when the study authors do not break apart fraternity and sorority membership. Use -9 for cannot tell.

[perathlete]
Enter the percent of participants in this group who reported being a member of a school athletic team. Use -9 for cannot tell.

[rape_myth]
Enter the mean rape myth acceptance score for participants in this group at baseline. Use -9 for cannot tell.

[empathy]
Enter the mean victim empathy score for participants in this group at baseline. Use -9 for cannot tell.

[per.sexual.victim]
Enter the percent of participants in this group who reported being a victim of sexual assault (by any type of perpetrator). Use -9 for cannot tell.

[per.physical.victim]
Enter the percent of participants in this group who reported being a victim of physical assault perpetrated by a partner (i.e., hookup, date, relationship). Use -9 for cannot tell.

[perknow]
Enter the percent of participants in this group who report knowing someone who has been a victim of IPV and/or sexual assault. Use -9 for cannot tell.

[efficacy_mean]
Enter the mean bystander efficacy score for participants in this group at baseline. Use -9 for cannot tell.

[perefficacy]
Enter the percent of participants in this group who reported confidence in their ability to intervene when witnessing violence or its warning signs (measure at baseline). Use -9 for cannot tell.

[intentions_mean]
Enter the mean bystander intention score for participants in this group at baseline. Use -9 for cannot tell.

**[perintentions]**
Enter the percent of participants in this group who reported intentions to intervene when witnessing violence or its warning signs (as measured at baseline). Use -9 for cannot tell.

**[bystanderbx_mean]**
Enter the mean number of times participants in this group reported engaging in pro-social bystander behavior (i.e., intervening when witnessing any type of violence or its warning signs) at baseline. Use -9 for cannot tell.

**[perbystanderbx]**
Enter the percent of participants in this group who reported engaging in pros-social bystander behavior (i.e., intervening when witnessing any type of violence or its warning signs) at baseline. Use -9 for cannot tell.

*Intervention group Characteristics*

**[bystander_tx]**
To what extent did the intervention focus on bystander behavior?

1. Bystander intervention was the main focus of the program.
2. General sexual assault prevention program with some bystander content. Note that sexual assault programs should not include content on physical violence (these will likely constitute IPV prevention programs).
3. General IPV prevention program with some bystander content. Note that IPV prevention programs can focus on physical violence exclusively or on both physical and sexual violence. Programs focusing exclusively on sexual violence should be coded as sexual assault prevention programs.
4. Other/Cannot tell. Please bring these to the attention of the PI, as they may not be eligible.

**[mixedsex]**
What was the sex composition of the intervention group (in delivery setting)?

1. All females.
2. All males.
5. Cannot tell.
6. NA – Individual Implementation.

**[sex_perp]**
How did the program conceptualize the sex of perpetrators of IPV/sexual assault?

1. Male perpetrators only.
2. Female perpetrators only.
3. Gender-neutral. Generic/ambiguous portrayal of sex or explicitly highlighting an equal possibility of male and female perpetrators.
4. Mostly male perpetrators.
5. Mostly female perpetrators.

**[sex_vic]**
How did the program conceptualize the sex of victims of IPV/sexual assault?
1. Male victims only.
2. Female victims only.
3. Gender-neutral. Generic/ambiguous portrayal of sex or explicitly highlighting an equal possibility of male and female victims.
4. Mostly male victims.
5. Mostly female victims.

**[gender_txt]**
Describe any intervention program content that addressed gender norms, rape myths, or other cultural norms as a root of IPV and/or sexual assault (or as an influence on bystander behavior). Where possible, please indicate whether gender/cultural norm content applies to sexual or physical violence. If a specific (branded) program is implemented (and a treatment manual or other program materials may be available) please clearly indicate this, as we may wish to obtain these materials for further coding.

**[format]**
In what format was the program typically delivered?
1. In-person. This involves some type of presentation or instruction delivered by one or more facilitators.
2. Video. This includes videos that are viewed on televisions or large screens as well as on computers (by groups or individual participants).
3. Web/computer delivered. This applies to non-interactive presentations that are viewed online as well as interactive computer-based programs. It excludes programs that primarily consist of participants viewing videos on computers (these should be coded as videos).
4. Ad/poster campaign. This applies to programs that are designed to promote awareness and contain minimal content (i.e., posters or flyers distributed across campus, brief emails sent to a listserv).
5. Other/cannot tell.

**[tx_context]**
In what context was the program typically delivered?
1. Individual with facilitator
2. Small groups (<10) with facilitator
3. Large group or whole classroom with facilitator
4. Individual alone (e.g., self-guided)
5. Other/cannot tell
[facilitator]
Who typically delivered the program?
1. Teachers or school administrators
2. Athletic coaches
3. Peer educators
4. Medical professionals (e.g., doctors, nurses, clinic staff)
5. Campus resource center staff (e.g., women’s center, sexual assault center, LGBTQ student center)
6. Non-campus community agency (e.g., domestic violence shelter staff/volunteers)
7. Mixed (no predominant type)
8. NA – Self-guided/implemented
9. Other/cannot tell

[monitored]
Monitoring of program implementation. Was the implementation of the program monitored by the author/researcher or program personnel to assess whether it was delivered as intended?
1. Yes, but no indication of feedback to facilitators. Do not infer that monitoring happened. Select “yes” ONLY if the report specifically indicates that implementation was monitored.
2. Yes, with indication that facilitators received feedback. Do not infer that feedback was provided. Select “yes” ONLY if the report specifically indicates that feedback was provided.
3. No indication that service delivery was monitored.

[impprob]
Implementation quality. Based on evidence or author acknowledgment, was there any uncontrolled variation or degradation in implementation or delivery of the program, e.g., high dropouts, erratic attendance, low compliance, program not delivered as intended, wide differences between settings or facilitators, etc. Note that this question has to do with variation in program delivery, not research contact. That is, there is no “dropout” if all participants complete treatment, even if some fail to complete the outcome measures.
1. Yes
2. Possible
3. No, apparently implemented as intended

[impfid_txt]
Implementation fidelity. Provide a description of any other implementation fidelity measures or assessments, including page numbers where appropriate.

[frequency_contact]
Approximate (or exact) frequency of contact between participant and facilitator or program activity. This refers only to the element of program that is different from what the control group receives.
1. Single-day program
2. Less than weekly
3. One to two times a week
4. Three to four times a week
5. Daily contact
6. Cannot tell

**[txwks]**
Duration of program in WEEKS. Approximate (or exact) mean number of weeks for the period over which participants received the program, from first to last intervention contact, excluding follow-ups designated as such. Divide days by 7; multiply months by 4.3; multiply years by 52; round to a whole number. Estimate for this item if necessary and if you can come up with a reasonable order of magnitude number (e.g., take the midpoint of a range if it is all that’s provided). Code -9 if cannot tell.

**[txhours]**
Duration of program as received in CONTACT HOURS. Approximate (or exact) mean number of contact hours for the period over which the participants received the program, from first program contact to last contact, excluding follow-ups designated as such. If dosage rates are provided multiply intended duration by dosage rate. Code -9 if cannot tell.

**[facilit]**
Facilitator training, preparation, or qualifications. Describe any information provided about the intervention facilitators’ training, level of preparation, or qualifications required for delivery. Provide page numbers.

**[incent]**
Incentives for recruitment or participation. Describe any incentives for participant recruitment and/or participation. Provide specific information about incentives (including dollar amounts), when available. Provide page numbers.
Outcomes

Step 1. Study and DV Identification
Create one record for each dependent variable that you will be coding. For example, if the study measures bystander intentions and bystander behavior, you will have two dependent variable records. This is different from the number of times a dependent variable is measured in a study. For example, if the study measures bystander intentions before and after the intervention, you will have only one record here – for the bystander intention measure (but you will have two effect sizes for this outcome measure: one at pretest and one at posttest).

[varid]
Variable number. This number is an identification number for the dependent variable you are coding. Each dependent variable is numbered consecutively, within the study you are coding so that each has a unique VarNo for that study. If there is only one dependent measure for this study, you will create only one record in this worksheet, and the variable number will be 1. If there are three dependent measures, they will be numbered 1, 2, and 3.

[dv.outcome.label]
Enter a label for the variable you are coding. When possible, use the exact label that the study report authors use.

[dvdes]
Write in a brief description of the measure you are coding. This should include the authors’ label for this variable (e.g., rape myth acceptance, bystander efficacy, etc.), the instrument, the direction of scoring (e.g., lower scores are better), and information about what is being measured. Quote or closely paraphrase the description that is provided in the original report.

For group equivalence variables make sure the label describes successes in terms of bystander attitudes and behaviors (e.g., females, younger participants, participants in lower education/grade levels, participants who have been IPV/sexual assault victims, participants who know an IPV/sexual assault victim, participants with lower rape myth acceptance or greater victim empathy). When coding race always default to white v. non-white. Code non-white as a success.

[dvmicro]
What type of measure are you coding?

Bystander Intervention Behavior
1. Intervention – sexual violence – female victim
2. Intervention – sexual violence – male victim
3. Intervention – sexual violence – unknown victim sex
4. Intervention – physical violence – female victim
5. Intervention – physical violence – male victim
6. Intervention – physical violence – unknown victim sex
7. Intervention – unclear violence type – female victim
8. Intervention – unclear violence type – male victim
9. Intervention – unclear violence type – unknown victim sex

**Bystander Attitudes**
10. Bystander efficacy
11. Bystander intentions

**Violence Perpetration**
12. Sexual Assault Perpetration – female victim
13. Sexual Assault Perpetration – male victim
14. Sexual Assault Perpetration – unknown victim sex
15. Physical IPV Perpetration – female victim
16. Physical IPV Perpetration – male victim
17. Physical IPV Perpetration – unknown victim sex
18. IPV Perpetration – unclear violence type – female victim
19. IPV Perpetration – unclear violence type – male victim
20. IPV Perpetration – unclear violence type – unknown victim sex

**Violence Attitudes**
21. Victim empathy
22. Rape myth acceptance

**Victimization**
23. Sexual Assault Victimization
24. Physical IPV Victimization
25. Know a victim of IPV or sexual assault

**Other Predictor Variables**
26. Age
27. Sex
28. Race/ethnicity
29. Grade/education level
30. Other predictor of bystander behavior or IPV/sexual assault

**Prerequisite Skills for Intervening**
31. Noticing a sexual assault or its warning signs
32. Identifying a situation as appropriate for intervention
33. Taking responsibility for acting/intervening
34. Knowing strategies for helping/intervening

**[dvtype]**
Type of data collection used for outcome measure.
1. Pencil & paper questionnaire
2. Online/computer assisted questionnaire
3. In-person interview
4. Phone interview
5. Other
6. Cannot tell

[dvdays]
Number of Days: Enter the number of days over which outcome was counted. Enter -8 for lifetime measures (e.g., “ever” experienced sexual assault). Enter -9 if cannot tell. Multiply months by 30 (e.g., enter 3 months as 90 days).

[icc]
For cluster randomized trials, please enter the intraclass correlation coefficient (ICC) for each outcome variable coded. Code -8 for not applicable and -9 for cannot tell.

Effect sizes

Although this is the final section of coding, it is a good idea to identify at least one codable effect size before you start coding a study, because studies that appear eligible frequently end up presenting data that cannot be coded into an effect size.

[reportid]
Report ID for this effect size. Indicate the report number (e.g., 22.02) for the report in which you found the information for this effect size. This is important so that we can find the source information for the effect sizes later on, if necessary, and is especially important for studies with multiple reports.

[page]
Page number for this effect size. Indicate the page number of the report identified above on which you found the effect size data. If you used data from two different pages, you can type in both, but use a comma or dash between the page numbers.

[estype]
Type of effect size you are coding.
1. Pretest and Posttest
2. Group equivalence

There are 3 types of effect sizes that can be coded: pretest, posttest, and group equivalence (or baseline similarity) effect sizes. They are defined as follows:

- **Group equivalence effect size.** Group equivalence effect sizes are used to code the equivalence of two groups prior to treatment delivery on (a) sex, (b) age, (c) race/ethnicity, or (d) predictors of pro-social bystander behavior. When multiple racial/ethnic group compositions are reported please report only white/non-white proportions (if not available, select another racial/ethnic group). When available, always code sex, age, and race/ethnicity.
When multiple other predictors are reported select the three deemed to be most relevant (NOTE that behaviors are more relevant than attitudes/intentions).

- **Pretest effect size.** This effect size measures the difference between an intervention and comparison group before the intervention (or at the beginning of the intervention) on the same variable used as an outcome measure. Note: Use pretest data for different analytic samples if available. (e.g., separate pretest data for different follow-up waves).

- **Posttest/follow-up effect size.** This effect size measures the difference between two groups after intervention receipt on some outcome variable. Some posttests can occur during the intervention (after intake), immediately after the intervention ends, or any subsequent follow-up period after the intervention ends.

**Group Selection**

[esgroup1] [esgroup2]
Select the groups being compared in this effect size. Always select the focal prevention program to be ‘group 1.’

**Dependent Variable Selection**

[varid]
Select the dependent variable for this effect size.

[estiming]
Timing of measurement. Approximate (or exact) number of weeks after the end of the intervention when measurement occurred. Divide days by 7; multiply months by 4.3. Enter -9 if cannot tell, but try to make an estimate if possible.

**Effect Size Calculation and Data Entry**

It is now time to identify the data you will use to calculate the effect size and to calculate the effect size yourself if necessary.

You need to determine what effect size format you will use for each effect size calculation. There are two general formats you can use, each with its own section in FileMaker:

1. Compute ES from means, sds, variances, test statistics, etc.
2. Compute ES from frequencies, proportions, contingency tables, odds, odds ratios, etc.

Also note that within each of the above effect size formats, effect sizes can be calculated from a variety of statistical estimates; to determine which data you should use for effect size calculation, please refer to the following guidelines in order of preference:

1. Compute ES from regression coefficients with statistical controls for pretest measures and other potential confounding measures at baseline.
2. Compute ES from univariate descriptive statistics (means, sds, frequencies, proportions).
3. Compute ES from test statistics (t, F, Chi square).
4. If significance tests statistics are unavailable or unusable but p-values and degrees of freedom (df) are available, determine the corresponding value of the test statistic (e.g., t,
chi-square) and compute ES as if that value had been reported. If you encounter these types of data, please see the PI.

Note that if the authors present both covariate adjusted and unadjusted means, you should use the covariate adjusted ones. If adjusted standard deviations are presented, however, they should not be used.

[esfavor]
Which group is favored?

For intervention-control comparisons, the intervention group is favored when it does “better” than the comparison group. The comparison group is favored when it does “better” than the intervention group.

For group equivalence comparisons, the intervention group is favored when it has a higher probability of pro-social bystander attitudes/behaviors than the comparison group (e.g., females, younger participants, participants in lower education/grade levels, participants who have been IPV/sexual assault victims, participants who know an IPV/sexual assault victim, participants with lower rape myth acceptance or greater victim empathy). For the time being, code non-white as a success.

Remember that you cannot rely on simple numerical values to determine which group is favored. For example, a researcher might assess rape myth acceptance, and report this as a score. A lower score is better than a higher score, so in this case a lower number, rather than a higher one, indicates a more favorable outcome.

Sometimes it may be difficult to tell which group is better off because a study uses multi-item measures in which it is unclear whether a high score or a low score is more favorable. In these situations, a thorough reading of the text from the results and discussions sections usually can bring to light the direction of effect – e.g., the authors will often state verbally which group did better on the measure you are coding, even when it is not clear in the data table.

Note that if you cannot determine which group has done better, you will not be able to calculate a numeric effect size. (You will still be able to create an effect size record—just not a numeric effect size.)

Select the group that has done “better”:
1. Intervention
2. Comparison
3. Neither, exactly equal
4. Cannot tell

[esdata]
Effect size derived from what type of statistics?
1. N successful/unsuccessful (frequencies)
2. Proportion successful/unsuccessful (percentage successful or not)
3. Means and SDs; means and variances; means and standard errors
4. Independent t-test
5. Chi-square statistic (1 degree of freedom)
6. Effect sizes as reported directly in the study
7. Other statistical approximation

[esadj]
For this effect size, did you use adjusted data (e.g., covariate adjusted means) or unadjusted data? If both unadjusted and adjusted data are presented, you should use the adjusted data for the group means or mean difference, but use unadjusted standard deviations or variances. Adjusted data are most frequently presented as part of an analysis of covariance (ANCOVA). The covariate is often either the pretest or some personal characteristic such as socioeconomic status.
   1. Unadjusted data
   2. Pretest adjusted data (or other baseline measure of an outcome variable construct)
   3. Data adjusted on some variable other than the pretest (e.g., socioeconomic status)
   4. Data adjusted on pretest plus some other variables

[estxasn]
Assigned N for the intervention group

[exctasn]
Assigned N for the comparison group

[estxobn]
Observed N for the intervention group

[exctobn]
Observed N for the comparison group

[estxmean]
Mean for intervention group

[esctmean]
Mean for comparison group

[estxsd]
Standard deviation for intervention group

[esctsd]
Standard deviation for comparison group

[escella]
N successful for intervention group

[escellc]
N successful for comparison group

\textbf{[escellb]}
N failed for intervention group

\textbf{[escelld]}
N failed for comparison group

\textbf{[esindt]}
Independent t-value

\textbf{[eschisq]}
$\chi^2$ (df=1)

\textbf{[esauth]}
Standardized mean difference (SMD) reported by authors

\textbf{[esor]}
Odds ratio (OR) reported by authors

\textbf{[eshdse]}
Standard error (for SMD or OR) reported by authors

\textit{Final Effect Size Determination}

\textbf{[es_fmsmd]}
Effect size value- standardized mean difference

\textbf{[es_fmor]}
Effect size value- odds ratio

Remember that you cannot rely on simple numerical values to determine which group has done better. For intervention-control comparisons, a positive effect size should indicate that the intervention group did “better” on the outcome measure than the comparison group, while a negative effect size indicates that the comparison group did “better” than the intervention group, and a zero effect size means that the two groups are exactly equal on the measure.

You must make sure that the sign of the effect size matches the way we think about direction, such that the effect size is positive when the intervention group (or posttest) is better and negative when the comparison group (or pretest) is better.

Effect sizes can range anywhere from around $-3$ to $+3$. However, you will most commonly see effect sizes in the $-1$ to $+1$ range.
Note: If the authors report an effect size, include that in your coding and use it for the final effect size value if no other information is reported. However, if the authors also include enough information to calculate the effect size, always calculate your own and report it in addition to that reported in the study.

[esprob]
Any problems coding this effect size?
About this review

Sexual assault is a significant problem among adolescents and college students across the world. One promising strategy for preventing these assaults is the implementation of bystander sexual assault prevention programs, which encourage young people to intervene when witnessing incidents or warning signs of sexual assault. This review examines the effects bystander programs have on knowledge and attitudes concerning sexual assault and bystander behavior, bystander intervention when witnessing sexual assault or its warning signs, and participants’ rates of perpetration of sexual assault.