Journal of Experimental Criminology (2022) 18:1–40
https://doi.org/10.1007/s11292-020-09442-x

Examining the effects of teen dating violence prevention programs: a systematic review and meta-analysis

Chelsey Lee[1] · Jennifer S. Wong[1]

Published online: 25 September 2020

Abstract

Objectives This meta-analysis examines the efficacy of programs at increasing knowledge about dating violence, changing attitudes, increasing bystander behaviors, and reducing incidents of dating violence perpetration and victimization.
Methods A systematic search yielded 38 studies contributing 73 independent effect sizes. Studies were pooled by outcome measure and ten moderators were used to examine the impacts of program and study characteristics.
Results Prevention programs had a significant, positive impact on measures of knowledge (ES = 0.566, $z = 3.59$), attitudes (ES = 0.191, $z = 3.88$), and violence perpetration (ES = .157, $z = 3.11$), but did not significantly impact experiences of victimization or bystander behaviors.
Conclusions Results indicate that dating violence prevention programs are effective at improving knowledge, attitudes, and some behaviors, providing support for the continued implementation of these programs with adolescent populations. Future research should investigate the impact of specific program content and long-term behavioral outcomes.

Keywords Dating violence · Adolescent · Teen · Prevention · Meta-analysis · Systematic review

Dating violence is a serious and prevalent issue among adolescents and young adults, with approximately 43% of police-reported victimization incidents in Canada occurring among youth aged 15–24 (Hotton Mahoney 2010). In the USA, an estimated 1 in 11

✉ Jennifer S. Wong
jenwong@sfu.ca

Chelsey Lee
chelseyl@sfu.ca

[1] School of Criminology, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

female and 1 in 15 male teens have experienced physical violence perpetrated by a dating partner (Centers for Disease Control and Prevention [CDC] 2019). Dating violence refers to the perpetration of physical, emotional, or sexual abuse, stalking, or verbal threats between two people in a dating relationship and has been associated with substantial negative consequences (Cascardi et al. 2018; CDC 2019; Hickman et al. 2004). Deleterious outcomes such as depression, anxiety, substance abuse, disordered eating, and physical injury have been associated with experiences of dating violence and have evidence of long-term impacts (Ackard and Neumark-Sztainer 2002; Callahan et al. 2003; Coker et al. 2000; Teten Tharp et al. 2017). Additionally, many adults who experience intimate partner violence first experienced such violence during an adolescent relationship (Hotton Mahoney 2010). Early intervention and prevention of dating violence behaviors is imperative.

Numerous adolescent dating violence prevention programs have been developed in recent years, with a heavy focus on school-based programs. Most are delivered in classrooms in middle or high schools and work to address dating violence by targeting negative attitudes and behaviors, and focusing on the development of healthy relationship and conflict resolution skills (e.g., Adler-Baeder et al. 2007; Lowe et al. 2008; MacGowan 1997). Findings across the many individual evaluations of dating violence prevention programs are mixed. Several meta-analytic reviews have also examined program effectiveness; however, these reviews present disparate results and each has an accompanying set of limitations. The current study seeks to expand and improve upon the existing meta-analytic literature regarding the impacts of adolescent dating violence prevention programs.

Prevalence of dating violence among adolescents

Adolescence is a critical point in human cognitive and personal development, which increases the risk of dating violence (Kerig 2010; Smith and Donnelly 2008). Numerous psychological changes occur as adolescents begin to build their sense of self and self-confidence (Lapierre et al. 2019); dating experiences that begin at this age are replete with new challenges which can lead to conflict and stress. These emotions can increase the likelihood of youth engaging in behaviors to maintain an intimate relationship, such as perpetrating or tolerating partner violence (Lapierre et al. 2019; Smith and Donnelly 2008). Further, adolescents often lack the knowledge and skills to properly address conflicts that may arise, which increases their risk for abusive behaviors (Kerig 2010; Smith and Donnelly 2008).

The prevalence of dating violence among adolescents is high across North American regional and national populations (Hickman et al. 2004). In a nationally representative survey of American youth aged 12–18, Taylor and Mumford (2016) found that 37% ($n = 667$) of the total sample of 1804 reported having been in a dating relationship. Among these youths, rates of partner violence were high, with 69% reporting victimization and 63% reporting perpetration of at least one type of physical, sexual, or psychological violence. Gender differences in the sample were minimal; 69% of both boys and girls had been victimized, while 62% of boys and 63% of girls had committed violence. These findings are consistent with rates found in other studies at both regional and local levels. For example, Niolon et al. (2015) surveyed middle school students in

four American cities; of the 1673 participants who had been in a dating relationship, more than three-quarters reported perpetrating at least one form of dating violence. Specific estimates across studies vary and often differ widely based on how dating violence is defined (i.e., emotional, physical, and/or sexual violence; Hoefer et al. 2015).

With respect to the prevalence of dating violence, rates of psychological violence are typically higher than rates of physical or sexual violence (Leen et al. 2013). Reported estimates of psychological perpetration range from 62 to 77% (Niolon et al. 2015; Taylor and Mumford 2016), while victimization rates range from 42 to 66% (Hickman et al. 2004; Taylor and Mumford 2016). Physical violence often has much lower estimates; rates of perpetration range from 12 to 33% and those of victimization span 8 to 25% (Hickman et al. 2004; Kann et al. 2018; Niolon et al. 2015; Taylor and Mumford 2016). Sexual dating violence is the least prevalent form, with perpetration estimates of 12 to15% and victimization rates of 5 to 18% (Hickman et al. 2004; Kann et al. 2018; Niolon et al. 2015; Taylor and Mumford 2016).

Consequences of dating violence

The high prevalence of dating violence is especially concerning due to the potential adverse consequences that can arise. Adolescence is a critical juncture in the development of personal identity and sense of self; as such, adolescents' overall psychological well-being and development are particularly susceptible to the negative impacts of dating violence (Callahan et al. 2003). Dating violence victimization is significantly related to increased experiences of post-traumatic stress, depression, and anxiety, the effects of which are known to be long lasting (Banyard and Cross 2008; Brown et al. 2009; Callahan et al. 2003; Foshee et al. 2013).

Dating violence victimization is also associated with numerous negative behavioral and physical effects. There is evidence that physical and sexual dating violence are associated with increased disordered eating, substance use, and risky sexual behaviors (Ackard and Neumark-Sztainer 2002; Coker et al. 2000; Silverman et al. 2001). Physical harm may also result; the extent of injuries sustained ranges from minor to severe. For example, Teten Tharp et al. (2017) found that 54% of youth in grades 8 through 12 who reported dating violence had sustained an injury as a result of the abuse, with physical injuries significantly more prevalent in females than in males.

Dating violence prevention programming

An important component of a comprehensive approach to addressing adolescent dating violence is prevention programming; the majority of programs ultimately seek to end dating violence and promote equitable and respectful relationships among youth (Crooks et al. 2019). These goals are typically approached through strategies intended to increase knowledge about dating violence and healthy/unhealthy relationships, modify attitudes and beliefs towards violence and aggression in the relationship context, increase positive behaviors and reduce/eliminate aggressive behaviors, and increase and develop the skills needed to create positive relationships (Crooks et al. 2019). Adolescent dating violence prevention programs are usually administered in-

person in middle and high schools, allowing program organizers to reach a large number of youths in an efficient manner (Temple et al. 2013). School-based programming also allows for early intervention, which is important for averting future violence (Black et al. 2017; Crooks et al. 2019).

Empirical evaluations of dating violence prevention programs have produced mixed findings. Generally, programs show positive results concerning increases in knowledge about and attitudes towards dating violence (e.g., Gardner et al. 2004), although the relationship is not always statistically significant. With respect to behavioral outcomes such as violence perpetration or victimization, findings vary considerably in both significance level and direction of effect, with studies producing both positive and negative impacts (e.g., Levesque et al. 2016; Wolfe et al. 2009). Given the variability in findings, meta-analysis is important for understanding the overall landscape of dating violence prevention program efficacy.

Prior meta-analytic reviews on dating violence prevention

Several systematic reviews of dating, intimate partner, and sexual violence prevention programs for adolescents have been conducted (e.g., Community Preventative Services Task Force (CPSTF), 2018; De Koker et al. 2014; DeGue et al. 2014; Graham et al. 2019); here, we focus specifically on the four previous meta-analytic reviews of school-based adolescent dating violence prevention programs (de la Rue et al. 2017; Edwards and Hinsz 2014; Fellmeth et al. 2015; and Ting 2009) (see Table 1). All four reviews focused on studies using two-group designs, the included programs focused on sexual violence as well as more general dating violence, and the included studies were published up until 2013 (at the latest). In terms of outcomes, the study by Ting (2009) included measures of knowledge and attitudes, the studies by de la Rue et al. (2017) and Fellmeth et al. (2015) included knowledge, attitudes, and behavioral measures, and Edwards and Hinsz (2014) used a combined attitudinal/behavioral measure of dating violence. Findings are inconsistent across the reviews, with de la Rue et al. (2017) and Ting (2009) reporting positive results regarding knowledge and attitudes, de la Rue et al. (2017) and Fellmeth et al. (2015) finding no treatment impact on behavioral outcomes, and the results from Edwards and Hinsz (2014) indicating an overall positive effect of dating violence prevention programs.

The current study aims to extend and improve upon the existing comprehensive reviews through several key differences. The first is in the definition of the primary programs of interest. The four existing meta-analyses all included some programs that solely target sexual violence (e.g., rape, sexual coercion, date/acquaintance rape); we contend that while sexual violence is one element of dating violence, it is only one of the many components. By focusing specifically (and only) on sexual violence, a program would fail to target the most common dating violence behaviors as it would exclude non-sexual physical and psychological violence. Further, although sexual violence can occur within a dating relationship, it can also occur in non-dating contexts (e.g., stranger, acquaintance rape). The current study limits the analysis to programs targeting general dating violence prevention and excludes programs solely targeting sexual violence.

Additionally, previous meta-analyses have limited the included studies to those with two-group designs and excluded single-group, pre-test–post-test designs. This is a

Table 1 Summary of previous meta-analyses

Study	Inclusion criteria	# studies	# studies that overlap with present analysis	Conclusions
de la Rue et al. (2017)	2-group randomized or quasi-experimental design; middle or high school age and setting; outcomes included knowledge of and attitudes towards dating violence and perpetration and victimization of physical, verbal, and sexual aggression; includes sexual violence programs, published prior to 2014	23	6	Positive effect on knowledge and attitudes; no effect on perpetration or victimization outcomes
Edwards and Hinsz (2014)	Randomized or quasi-experimental design; middle and high school age; outcomes included attitudes towards sexual and dating violence, self- and school-reported physical and sexual violence perpetration; includes sexual violence programs; published prior to 2012	7	1	Positive effect on a combined measure of dating violence outcomes
Fellmeth et al. (2015)	Randomized or quasi-experimental design; middle, high school, and college programs; outcomes included victimization of dating violence, improvement in mental well-being, and improvements in behavior and knowledge about dating violence; includes sexual violence programs; published prior to 2013	33	2	No significant effects on knowledge, attitudes, behaviors, or skills
Ting (2009)	2-group designs; middle and high school age; outcomes included knowledge and attitudes towards dating violence; includes sexual violence programs; limited to peer-reviewed sources and official reports; published 1990–2007	13	3	Positive effect on knowledge and attitudes

common criterion in meta-analysis to ensure the quality of included studies is strong; designs without a control or matched comparison group may overestimate the treatment effect (Carlson and Schmidt 1999; Lipsey and Wilson 2001). However, due to logistical constraints when dealing with adolescent, school-based populations, randomized control trials or designs with matched comparison groups are often not possible. Our literature search determined that implementation of this inclusion criterion would eliminate a considerable number of evaluations of dating violence prevention programs from the analysis. By excluding those studies with weaker research designs, the conclusions that can be drawn from the field are arguably biased as many existing program evaluations are ignored. While there are limitations to single-group designs, some methodological techniques can be used to mitigate these concerns (see Morris

and DeShon 2002), and we contend that the inclusion of well-implemented single-group, pre-test–post-test designs allows for a more thorough and comprehensive examination of the field.

The present study also differs from the Ting (2009) meta-analysis in that we include unpublished works in addition to peer-reviewed, published studies (e.g., theses, dissertations, private research reports). Doing so provides a more comprehensive view of the field and minimizes potential publication bias. Finally, the current meta-analysis provides an up-to-date examination of dating violence prevention programs (initial search conducted in April 2019 and updated in February 2020), a timeframe which extends beyond prior reviews by a minimum of 6 years. By using a more current literature search, a more specific definition of dating violence prevention, and including additional research designs and unpublished works, the current study has very little overlap with previous meta-analyses on dating violence (only 6 of our 38 included studies overlapped with the existing reviews). These distinctions help the present study to provide important insights into the effectiveness of adolescent dating violence prevention programs.

Study aim

The purpose of the study is to use systematic review and meta-analysis to synthesize and evaluate the overall effectiveness of dating violence prevention programs at improving adolescent knowledge, attitudes, and behaviors. Specifically, this study examines the impact of dating violence prevention programs at (1) increasing knowledge about dating violence, (2) improving attitudes towards dating violence behaviors, (3) reducing incidents of dating violence perpetration, (4) reducing experiences of dating violence victimization, and (5) increasing bystander intentions and/or behaviors.

Method

Systematic review strategy for identification of studies

A comprehensive and exhaustive list of terms was first identified by consulting the relevant literature for common terms as well as synonyms and interchangeable phrases. The constructs were developed using an iterative process of trial and error, with multiple rounds of testing. Boolean operators and wildcard markers were used to broaden the search. The final set of search terms involved three constructs, as shown in Table 2.

The final iteration of the search strategy was applied to 23 electronic databases (e.g., Criminal Justice Abstracts, PsycINFO; see Appendix for full list) and the key terms were searched in the Abstract field. Additionally, in an effort to avoid issues of publication bias, the grey literature was searched for unpublished works such as theses and dissertations, conference presentations, technical reports, and independent projects. Grey literature sources included Google and Google Scholar (the first 100 hits of each were searched), websites of relevant organizations (e.g., Public Safety Canada, Department of Justice, National Resource Center on Domestic Violence), and the curriculum vitae of prominent researchers in the field (e.g., Vangie Foshee, Bruce Taylor). The reference lists of included studies were also hand searched to identify additional studies,

Table 2 Summary of search constructs and terms

Construct	Search terms
Dating violence	"dating violence" OR "relationship violence" OR "dating aggression" OR "relationship aggression" OR "couple violence" OR "dating relationship*" OR "healthy relationships" OR "dating abuse" OR "relationship abuse" OR "dating matters" OR "ending violence" OR "strong start" OR "expect respect" OR "teen choices" OR "safe dates" OR "date SMART" OR "shifting boundaries" OR "Courage7&8" OR "The fourth R" OR "A.S.A.P. "OR "SafeTeen" OR "Voices Against Violence"
Prevention program	program OR prevent* OR interven* OR campaign
Evaluation	eval* OR impact* OR outcome* OR assess* OR effect*

as were the reference lists of identified meta-analyses or literature reviews focused on dating violence prevention or relationship education programs.

Selection criteria

Inclusion criteria To be included in the pooled analysis, (1) studies must have evaluated a program specifically focused on dating violence prevention or education for an adolescent population (under age 18). (2) The study must have included at least one quantitative outcome measure relevant to knowledge about dating violence, attitudes towards dating violence, dating violence victimization or perpetration, or bystander behaviors. (3) The study must also have provided sufficient data and information to enable the calculation of an effect size. (4) Included studies used a control/comparison group and/or a pre-test/post-test design and scored at least a two on the Maryland Scale of Scientific Methods (i.e., "temporal sequence between the program and the crime or risk outcome clearly observed, or the presence of a comparison group without demonstrated comparability to the treatment group"; Sherman et al. 1998, p. 4). This included randomized control trials, quasi-experimental studies (with a matched comparison group or wait-list control group), and single-group, pre-test–post-test studies. Studies were eligible for inclusion if they (5) used a sample size of at least 20 participants, (6) were published in English or French, and (7) were conducted in North America, Western Europe, Australia, or New Zealand.

Exclusion criteria To further refine the sample and ensure commensurability among studies, a set of exclusion criteria were applied. Studies were excluded if the primary focus was the prevention of any other type of violence not identified as dating violence and that does not occur within a specified dating relationship. This included studies focused solely on sexual violence, general school violence, non-intimate partner peer violence, and bullying. Studies were also excluded if the target population was very specific (e.g., those with particular mental health concerns or histories of serious maltreatment or abuse, teen mothers, youths living in a residential facility) to maximize similarities across study populations. In addition, studies were excluded if the research design was a post-test only design, or if the comparison group was deemed inappropriate (e.g., those exposed to another program or intervention).

Study collection

Once the initial literature search was performed and studies were identified as being potentially relevant, two independent reviewers read through the titles and abstracts and applied inclusion criteria to determine those studies that should be retrieved for further review. After retrieval, two reviewers applied the inclusion and exclusion criteria to identify those to be retained for coding. Discrepancies in reviewer decisions were discussed until an agreement was reached, leading to a final sample of 38 studies.

Analytic approach

Coding All studies selected for inclusion were coded independently by two reviewers on a series of 83 variables. These variables included program characteristics (e.g., program goals, program components), intervention characteristics (e.g., number of sessions and contact hours, who delivered the program material), study characteristics (e.g., research design, type of comparison group; score on the Maryland Scale of Scientific Methods), sample characteristics (e.g., sample size, age range of participants, gender), outcome measures (e.g., type of outcome, type of data, timing of post-test), and treatment and control group outcomes (e.g., mean scores on measurement scales at pre-test and post-test).

Effect size calculation Due to the differing forms of data and statistical methods used across studies, it is necessary to transform individual study findings into a standardized format to conduct the pooled analysis (Card 2011; Hedges 2000). We used a standardized mean difference as the primary effect size type; calculations were based on the following types of reported outcome data: (a) For the 16 studies that presented means and standard deviations, a Cohen's *d* was calculated. (b) For studies that did not report means and/or standard deviations, a standardized mean difference was computed using other available statistics including a beta coefficient and standard deviation ($n = 1$), and an F test with unequal group sizes ($n = 1$). (c) Nine of the included studies presented dichotomous outcome measures (e.g., percentages or raw numbers representing how many participants experienced or perpetrated dating violence). Effect sizes for these studies were computed as odds ratios (Lipsey and Wilson 2001), and we applied a Cox transformation to the logged odds ratios to adjust for the dichotomous data (Lipsey and Wilson 2001; Sanchez-Meca et al. 2003). (d) Last, 11 studies used a single-group pre-test–post-test design.[1] These effect sizes were calculated as an adapted Cohen's *d* using

[1] It is widely debated whether effect sizes from single-group designs can be appropriately combined with those from two-group designs (Cuijpers et al. 2017; Lipsey and Wilson 2001). Some researchers have argued that effect sizes can be pooled if they provide estimates of the same treatment effect and are computed in the same metric (Morris and DeShon 2002). To determine whether there is appropriate equivalence between studies, researchers must first consider how well a single-group study controls for potential bias and other factors that could impact the reliability of the treatment estimate (Borenstein and Hedges 2019; Morris and DeShon 2002). Additionally, as effect sizes between single-group and two-group designs are inherently different (due to using different standard deviations; e.g., standard deviations of pretest scores versus post-test scores versus pooled scores), the scaling of the effect size will differ and must be transformed to correct for disparities (Morris and DeShon 2002). Following these suggestions, the single-group evaluations in the current study were assessed for comparability to the treatment effect in the two-group studies, and effect sizes were computed using an adapted Cohen's *d* and transformed to a raw score metric.

pre-test and post-test means and standard deviations, then transformed from a change score metric to a raw score metric to be commensurable with the effect sizes produced from two-group designs (see Morris and DeShon 2002).

Cluster adjustments Many of the included studies did not assign individual students to the treatment or comparison groups; instead, conditions were assigned based on groups such as whole schools or classrooms within schools. As participants in these designs were nested within clusters, we cluster-adjusted the effect sizes and standard errors (Hedges and Hedberg 2007). These adjustments were made using an estimate of the intra-cluster correlation coefficient (ICC), a measure of the proportional variance attributable to group differences. Of the 38 studies included, only seven reported an exact ICC or a range of ICCs across outcomes. For the remaining studies, we turned to the literature to determine an appropriate estimate of the ICC for each cluster type. Unfortunately, guidelines with respect to ICCs are lacking, with sparse information available on appropriate ICC estimates for similar cluster (classroom, school) and outcome types (behavior, attitudes). Based on the reported ICCs in our sample and the existing literature (including the What Works Clearinghouse (2017) along with several primary studies: Brackett et al. 2011; Di Stasio et al. 2016; Fonagy et al. 2009; Isaacs et al. 2013; Murray and Blitstein 2003; Spence et al. 2003; Thornberg et al. 2017), we estimated the most appropriate ICC estimates for the current set of studies. We used an ICC of .026 for behavioral outcomes and an ICC of .028 for attitudinal outcomes for the school-level cluster adjustments, and an ICC of .10 for classroom-level cluster adjustments.

Pooled effect estimates There are two primary models used in meta-analysis: fixed effects and random effects models. Both models weight each study by its inverse variance, however, the calculation of the weights differs based on the assumed source of variability between studies (Card 2011; Egger and Smith 2001; Lipsey and Wilson 2001). A fixed effects model assumes the between-subject variability present is the result of sampling error and occurs only by chance (Lipsey and Wilson 2001), while random effects models assume that between-study heterogeneity is significant and variability is due to factors other than random subject-level sampling error (Card 2011; Lipsey and Wilson 2001). In the current analysis, given a multitude of between-study differences and the fact that the specific dating violence programs varied across studies, we implemented random effects models. Publication bias was assessed using Egger's test of small study effects and funnel plots (Steichen 1998; Sterne and Harbord 2004). Influence analysis was used to identify potential outliers, and heterogeneity was assessed using Q-statistics and I^2 statistics. The Q-statistic indicates if there is significant heterogeneity between studies, while the I^2 statistic is a measure of how much heterogeneity can be attributed to sources other than sampling error (Card 2011; Lipsey and Wilson 2001).

Outcome measures The present analysis includes a series of five outcome measures; multiple effect sizes were included per individual study as long as each effect size was pooled within separate outcome measure categories (i.e., a given study could not contribute more than one effect size per category). The five outcomes are the following: (1) knowledge of dating violence (e.g., recognition of types of dating violence

behaviors or facts about dating violence), (2) attitudes towards dating violence (e.g., beliefs about if/when violence is appropriate, acceptance of rape myths), (3) perpetration of dating violence behaviors (e.g., hitting, slapping, name-calling), (4) experiences of dating violence victimization (e.g., being hit or verbally abused by a dating partner), and (5) bystander behaviors and intentions (e.g., "if your friend called their partner a rude name, how likely is it that you would say something?").

Subgroup analysis Additionally, subgroup analysis was used to assess potential sources of variability among various program and study characteristics (Lipsey and Wilson 2001). A series of ten dichotomous variables were selected as potential moderating variables including (1) length of program (≤ 5 h vs. > 5 h); (2) curriculum addressed gender roles and stereotypes (yes/no); (3) curriculum incorporated bystander training (yes/no); (4) level of school (middle school vs. high school); (5) treatment group sample size (< 300 participants vs. ≥ 300 participants); (6) randomized research design (randomized control trial vs. non-randomized design); (7) single-group research design (single-group pre-test–post-test design vs. two-group design); (8) publication year (prior to 2015 vs. 2015 or later); (9) time of post-test (immediately after program end vs. 1+ months after program end); (10) sample ethnicity (group was predominantly Caucasian or mixed ethnicity vs. predominantly ethnic minority).[2] Subgroup analysis was conducted using the analog to the ANOVA method, which separates the total variability (Q_T) into the within-group variation (Q_w = the summed Q-statistics for each of the two groups in the analysis) and that which can be explained by the categorical variable (the between-group variation; Q_b = the difference between the total and within Q-statistics). If the Q_b is statistically significant, it suggests the two categories are producing significantly different effect sizes and the difference is due to more than sampling error (Lipsey and Wilson 2001).

Decision rules A key assumption in primary data analysis is the independence of observations; this assumption holds true for meta-analysis with effect size as the unit of observation (Card 2011). To ensure independence of effect sizes, several decision rules were implemented: (a) If multiple documents presented data from the same population or study, the one with the most detailed information and relevant outcomes was selected. (b) If a study separated their sample into independent subgroups (e.g., by gender), results were combined across subgroups to create a single effect size. (c) If a study included multiple treatment groups, the treatment group that was the most comparable to those in the overall set of studies was selected for inclusion. (d) When multiple post-tests were reported (e.g., an immediate post-test and a 3-month follow-up), the immediate post-test was chosen to maximize comparability across all included studies (as this was the most common time point reported). (e) Some studies reported multiple outcomes that were categorized under the same outcome measure category (e.g., knowledge of dating violence); in these situations, the most commensurate outcome to other studies was selected. (f) When a study reported multiple measures

[2] Due to inconsistencies in reporting and differing sample sizes between outcomes, not all moderating variables could be used for all outcomes. The specific moderating variables used in each analysis are identified in the corresponding section for each of the five outcomes (dating violence knowledge, attitudes towards dating violence, perpetration of violence, victimization of dating violence, and bystander behaviors).

of violence perpetration/victimization using the same sample, they were prioritized in the order of emotional, then physical, then sexual violence. Emotional violence was prioritized as this form of violence is reported with high frequency among youth (Leen et al. 2013; Taylor and Mumford 2016).

Throughout the coding process, "missing" data was often noted; in other words, across the set of included studies, certain pieces of information may not have been presented by study authors (e.g., sample size, participant age, whether the program included bystander training). We attempted to obtain this information by contacting study authors; otherwise, when possible, we inferred the missing data or calculated it using data available in the evaluation report.[3] Notably, there was substantial inconsistency across studies in reporting whether certain components were included in the program curricula. In some cases, it was possible to infer the information based on other information provided; for example, it was assumed that group discussion was involved if the program was delivered in-person in a small group setting.

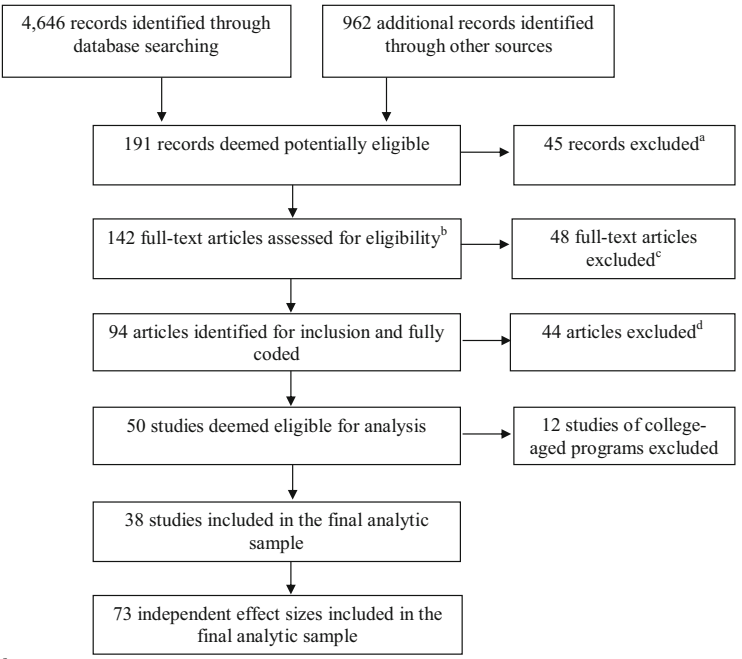
Results

The initial search of the 23 electronic databases resulted in a total of 4646 articles, while the search of the grey literature identified 962 additional studies (see Fig. 1). The abstracts of all hits were reviewed and 146 articles were identified as potentially eligible; inclusion and exclusion criteria were applied and 94 studies were selected. Interrater reliability was high for retrieval agreement with a kappa of 0.783 ($z = 10.82$, $p < .001$) and moderately high 0.621 ($z = 7.30$, $p < .001$) for inclusion agreement. All disagreements were discussed and resolved. Due to reasons such as missing data, data that were inappropriate for calculating commensurable effect sizes, or overlapping samples (i.e., studies using the same dataset), an additional 44 studies were excluded during the data extraction and effect size calculation stages (remaining $n = 50$). Upon consideration of clear programmatic heterogeneity based on participant age, programs targeting middle/high school students and college students were deemed incommensurate for pooling. College programs were excluded, leaving a final sample of 38 independent studies which contributed 73 independent effect sizes (see Table 3).

Overview of the included studies

The 73 effect sizes were dispersed across the five dating violence outcome measures: knowledge (22%), attitudes (27%), perpetration (23%), victimization (18%), and

[3] For example, several studies reported only the total or combined pre- or post-test sample sizes, rather than specifying the sample sizes per treatment/control group. In these cases, when other information regarding the sample size was available such as degrees of freedom or full pre- or post-test sample size, we calculated the group sample sizes using an assumption of proportional attrition. For example, we took the analytic sample size per effect size based on the degrees of freedom from the reported F tests and assumed proportionally equivalent attrition from the treatment and control groups based on pre-test sample sizes.



[a] After the initial set of 191 records was identified, two reviewers assessed the abstracts for relevance and excluded 45 records that were deemed to be outside the scope of the study.

[b] Four additional sources were identified, but we were unable to obtain the full-text.

[c] During the application of inclusion/exclusion criteria, a total of 48 studies were excluded for the following reasons: 15 studies were excluded due to data inappropriate for calculating effect sizes or a lack of quantitative data; 6 were excluded due to irrelevant outcomes (e.g., communication strategies, knowledge of general, non-dating violence behaviors); 3 were excluded due to overlapping samples; 5 were excluded due to an inappropriate comparison group (e.g., another program); 1 study was excluded due to small sample size; and the remaining 18 studies were excluded due to other reasons (e.g., not being an evaluation, published in a non-Western country, not published in English, targeted too specific of a population, or was a post-test only design).

[d] During the data extraction and coding processes, an additional 44 studies were excluded for the following reasons: 29 studies were excluded due to missing data or an inability to calculate effect sizes (e.g., missing sample sizes, only reporting post-test data, reporting a statistical test not amenable to calculating standardized mean difference effect sizes); 4 were excluded due to overlapping samples; 1 was excluded for methodological issues (i.e., biased sample, imputed outcome data); 6 were excluded for incommensurate outcome measures (e.g., likelihood to engage in stages of change in the context of bystander behavior; anger management, not reporting summary results, only by single questions); 2 were excluded due to small sample size; and 2 were excluded as a result of an inappropriate control group (e.g., another program or condition of program).

Fig. 1 Systematic search results

bystander behaviors and intentions (10%). Studies were published between 1997 and 2019 and a majority (82%) were published in peer-reviewed journals (see Table 4). Over half of the included studies used a randomized control trial design (58%) and half used an immediate post-test (50%). Treatment group sample size ranged from 20 to 1389 participants at pre-test, and sample size at post-test ranged from 20 to 1267. Participant mean age was 14.46 (SD = 1.2). Of the studies reporting the racial/ethnic composition of the sample ($n = 32$), just over one-third used a predominantly Caucasian or mixed ethnicity sample (34%), while 66% involved a sample that primarily included ethnic minority participants. The 38 programs were mostly delivered in a classroom environment (74%), and program length ranged from 15 min to 28 h.

Table 3 Study and program characteristics and outcomes

Author (date)	Program name	Research design (Maryland Scale Rating)	Sample characteristics (mean age; gender distribution; ethnic composition)	Sample size and composition		Program delivery (setting; facilitator)
				Treatment group	Control group	
Adler-Baeder et al. (2007)	Love U2: Increasing Your Relationship Smarts (RS adapted)	2[a]	Mean age 16.1; mixed female/male; mixed ethnicity	235 students from 9 classes across 9 schools	105 students from 9 classes across 9 schools[h]	Classroom; school teachers
Antle et al. (2011)	Love U2: Communication Smarts	2	Mixed female/male; predominant minority[e]	202 youth involved with a specialized program in public schools	N/A	Classroom; specialized Y.O.U. school staff
Avery-Leaf et al. (1997)	Not specified	5[d]	Mean age 16.5; mixed female/male; Caucasian	102 students from 3 health classes	90 students from 3 health classes	Classroom; school health teachers
Ball et al. (2012)	Expect Respect	2	Mean age 14; mixed female/male; predominant minority	144 students who previously reported experiences of violence	N/A	Classroom; program facilitators
Banyard et al. (2019)	Reducing Sexism and Violence Program-Middle School Program (RSVP-MSP)	4[c]	Mean age 12.5; male only [f]	211 students from 3 public schools	129 students from 1 public school	Classroom; program facilitators
Belknap et al. (2013)	Not specified	2	Mean age 13.4; mixed female/male; predominant minority	66 students from 3 middles schools	N/A	Small group; undergraduate students
Chamberland et al. (2014)	Session d'Ateliers Interactifs de Sensibilisation, d'Information et de Réflexion à la violence dans les relations amoureuses des adolescentes (SAISIR)	5	Mean age 14.7; mixed female/male[f]	484 students from 5 high schools	287 students from 4 high schools	Classroom; women's shelter staff

Table 3 (continued)

Author (date)	Program name	Research design (Maryland Scale Rating)	Sample characteristics (mean age; gender distribution; ethnic composition)	Sample size and composition		Program delivery (setting; facilitator)
				Treatment group	Control group	
Cissner and Hassoun Ayoub (2014)	Fourth R	5	Mean age 12; mixed female/male; predominant minority	307 students from 24 classes across 10 schools	263 students from 23 classes across 10 schools	Classroom; teachers
Connolly et al. (2014)	Respect in Schools Everywhere (RISE)	5	Mean age 12.37; mixed female/male; predominant minority	209 students from 2 schools	300 students from 2 schools	Classroom; high school students and mental health workers
de Graaf et al. (2015)	Rock and Water	4	Mean age 15.2; male only; Caucasian	260 students from 4 schools	261 students from 4 schools	Classroom; phys ed. teachers
DeGannes (2009)	Project Awareness	2	Mean age 13.46; mixed female/male; predominant minority	57 middle school students	N/A	Classroom; certified rape crisis counselors
Edwards et al. (2019)	Bringing in the Bystander–High School Curriculum	5	Mean age 15.8; mixed female/male; Caucasian	1081 high school students from 11 schools	1322 high school students from 14 schools	Classroom; program facilitators
Gardner et al. (2004)	Connections: Relationships and Marriage	3	Mean age 16.5; mixed female/male; predominant minority	263 students from 6 classes	147 students from 6 classes	Classroom; teachers
Gonzalez-Guarda et al. (2015)	JOVEN/YOUTH (Juntos Opuestos a la Violence Entre Novios/Together Against Dating Violence)	5	Mean age 14.34; mixed female/male; predominant minority	41 student/ parent dyads	41 student/parent dyads	Small group; program facilitators
Jaycox et al. (2006)	Ending Violence: A curriculum for educating teens on domestic violence and the law	5	Mean age 14.41; mixed female/male; predominant minority	1240 students from 55 classes	1057 students from 55 classes	Classroom; attorneys

Table 3 (continued)

Author (date)	Program name	Research design (Maryland Scale Rating)	Sample characteristics (mean age; gender distribution; ethnic composition)	Sample size and composition		Program delivery (setting; facilitator)
				Treatment group	Control group	
Joppa et al. (2016)	Katie Brown Educational Program	5	Mean age 15.85; mixed female/male; Caucasian	86 students from 11 classes	139 students from 13 classes	Classroom; program facilitators
Jouriles et al. (2019)	TakeCare	5	Mean age 15.7; mixed female/male; predominant minority	85 high school students	80 high school students	Small group; video facilitated by research staff
Levesque et al. (2016)	Teen Choices	5	Mean age 15; mixed female/male; Caucasian	1752 students from 10 schools	1690 students from 10 schools	Online; delivered online
Lowe et al. 2008)	Safe Relationships Program	2	Mean age 15.4; mixed female/male; mixed ethnicity	106 students participating in health class	N/A	Classroom; staff from involved agencies
MacGowan (1997)	Not specified	5	Mean age 12.6; mixed female/male; predominant minority	241 students from 8 classes	199 students from 7 classes	Classroom; teachers
McLeod et al. (2015)	The Chesterfield RELATE (Relationship Education Leading Adolescents towards Empowerment) program	2	Mean age 14.5; mixed female/male; predominant minority	279 high school students	N/A	Classroom; peer facilitators
Miller (1998)	Teen Dating Violence Intervention and Prevention	5	Mean age 12.8; mixed female/male; predominant minority	20 students with a history of violence	21 students with a history of violence	Small group counseling; counselors
Miller et al. (2012)	Coaching Boys into Men	5	Mean age 15.5; male only; predominant minority	1008 student athletes from 8 schools	998 student athletes from 8 schools	Small group; coaches
Miller et al. (2015)	SHARP	5	Mean age 15.5; mixed female/male; predominant minority	495 students from 4 schools	516 students from 3 schools	School health center; health center staff

Table 3 (continued)

Author (date)	Program name	Research design (Maryland Scale Rating)	Sample characteristics (mean age; gender distribution; ethnic composition)	Sample size and composition		Program delivery (setting; facilitator)
				Treatment group	Control group	
Munoz-Fernandez et al. (2019)	Dat-e Adolescence	5	Mean age 14.98; mixed female/male[f]	557 students from 4 high schools	866 students from 3 high schools	Classroom; researchers and peers
Peskin et al. (2014)	It's Your Game…Keep it Real	5	Mean age 13; mixed female/male; predominant minority	598 students from 5 schools	758 students from 5 schools	Classroom; program facilitators
Peskin et al. (2019)	Me & You: Building Healthy Relationships	5	Mean age 12.4; mixed female/male; predominant minority	192 students from 5 middle schools	151 students from 5 middle schools	Classroom and online; trained facilitators
Reidy et al. (2017)	Expect Respect Support Groups	4	Mean age 14.3; mixed female/male; predominant minority	901 students from 24 schools	777 students from 12 schools	Classroom; program facilitators
Rizzo and Houck (2019)	STRONG	5	Mean age 13; male only; Caucasian	59 parent-son pairs	60 parent-son pairs	Online; online
Rizzo et al. (2018)	Project Date Smart	5	Mean age 15.75; female only; predominant minority	59 adolescents with a history of dating violence	50 adolescents with a history of dating violence	Classroom; program facilitators
Sanchez-Cesareo (2002)	Project Youth	3	Mean age 14.6; mixed female/male; predominant minority	610 students from 2 high schools	227 students from 1 high school	Classroom; college student facilitators
Sanchez-Jimenez et al. (2018)	Dat-e Adolescence	5	Mean age 14.73; mixed female/male[f]	908 students from 3 high schools	856 students from 4 high schools	Classroom; researchers and peers
Sargent et al. (2017)	TakeCare	5	Mean age 15.27; mixed female/male; predominant minority	463 students from 29 classes	458 students from 28 classes	Classroom; school counselors

Table 3 (continued)

Author (date)	Program name	Research design (Maryland Scale Rating)	Sample characteristics (mean age; gender distribution; ethnic composition)	Sample size and composition		Program delivery (setting; facilitator)
				Treatment group	Control group	
Schramm and Gomez-Scott (2012)	Relationship Smarts Plus (RS+)	2	Mean age 15; mixed female/male; Caucasian	426 students from 22 classes	197 students from 22 classes[i]	Classroom; teachers
Silverman (1999)	Teen Dating Violence Prevention and Intervention	2	Mean age 14.8; mixed female/male; Caucasian	199 students from 2 schools	N/A	Classroom; school personnel
Southgate (2016)	Not specified	2	Mean age 14.4; female only; mixed ethnicity	30 adolescents	N/A	Small group[g]
Wolfe et al. (2009)	Fourth R: Skills for Youth Relationships	5	Mean age 14.5; mixed female/male[f]	968 students from 10 schools	754 students from 10 schools	Classroom; teachers

Author (date)	Discussed gender roles	Included bystander training	Program length	Outcomes	Effect size (standard error)
Adler-Baeder et al. (2007)	No	No	15 h over 12 sessions	Attitudes	0.079 (0.143)
				Violence perpetration	0.271 (0.146)
				Knowledge	1.696 (0.113)
Antle et al. (2011)	Missing	Missing	8 h over 2 sessions	Attitudes	0.125 (0.076)
				Knowledge	0.914 (0.087)
Avery-Leaf et al. (1997)	Yes	Missing	5 h over 5 sessions	Attitudes	0.256 (0.290)
Ball et al. (2012)	Yes	Missing	22 h over 24 session	Violence victimization	0.032 (0.083)
				Violence perpetration	0.076 (0.119)
Banyard et al. (2019)	Yes	Yes	4 h over 4 sessions	Bystander behavior	− 0.067 (0.287)
				Attitudes	0.352 (0.288)

Table 3 (continued)

Author (date)	Discussed gender roles	Included bystander training	Program length	Outcomes	Effect size (standard error)
Belknap et al. (2013)	No	No	1.5 h over 1 session	Attitudes	0.107 (0.081)
Chamberland et al. (2014)	Yes	No	5 h over 4 sessions	Attitudes	0.332 (0.140)
				Knowledge	0.395 (0.141)
Cissner and Hassoun Ayoub (2014)	Yes	Yes	26 h over 21 sessions	Violence victimization	− 0.160 (0.200)
				Violence perpetration	− 0.124 (0.170)
Connolly et al. (2014)	No	Missing	1.5 h over 2 sessions	Knowledge	0.257 (0.190)
				Violence victimization	0.655 (0.186)
				Attitudes	− 0.039 (0.190)
de Graaf et al. (2015)	Missing	No	10 h over 10 sessions	Attitudes	0.467 (0.147)
				Violence perpetration	0.406 (0.174)
DeGannes (2009)	No	No	8 h over 8 sessions	Attitudes	0.417 (0.208)
				Knowledge	0.403 (0.219)
Edwards et al. (2019)	No	Yes	5.5 h over 7 sessions	Violence perpetration	0.064 (0.130)
				Violence victimization	0.073 (0.108)
Gardner et al. (2004)	No	No	15 h over 15 sessions	Violence perpetration	0.174 (0.138)
				Knowledge	0.564 (0.142)
Gonzalez-Guarda et al. (2015)	Yes	Yes	6 h over 6 sessions	Violence perpetration	0.511 (0.272)
				Violence victimization	0.456 (0.296)
Jaycox et al. (2006)	No	No	3 h over 3 sessions	Knowledge	0.706 (0.073)
				Violence victimization	− 0.020 (0.088)
				Attitudes	− 0.030 (0.072)

Table 3 (continued)

Author (date)	Discussed gender roles	Included bystander training	Program length	Outcomes	Effect size (standard error)
Joppa et al. (2016)	Yes	No	5 h over 5 sessions	Attitudes	0.539 (0.184)
				Knowledge	− 0.377 (0.183)
Jouriles et al. (2019)	No	Yes	0.5 h over 1 session	Bystander behavior	0.457 (0.158)
Levesque et al. (2016)	No	No	1.5 h over 3 sessions	Violence perpetration	0.475 (0.073)
				Violence victimization	0.419 (0.076)
				Attitudes	0.141 (0.074)
Lowe et al. 2008)	No	No	4 h over 4 sessions	Attitudes	0.062 (0.090)
				Knowledge	0.402 (0.116)
MacGowan (1997)	No	Missing	5 h over 5 sessions	Knowledge	0.030 (0.187)
McLeod et al. (2015)	No	Yes	5 h over 5 sessions	Attitudes	0.946 (0.159)
				Knowledge	1.125 (0.113)
Miller (1998)	Yes	No	10 h over 10 sessions	Attitudes	− 0.041 (0.312)
Miller et al. (2012)	Missing	Yes	3.75 h over 11 sessions	Violence perpetration	0.102 (0.104)
				Attitudes	0.000 (0.094)
				Bystander behavior	0.182 (0.104)
				Knowledge	0.059 (0.102)
Miller et al. (2015)	No	Yes	15 min over 1 session	Bystander behavior	0.060 (0.159)
				Knowledge	0.135 (0.136)
Munoz-Fernandez et al. (2019)	No	Yes	7 h over 7 session	Violence perpetration	0.032 (0.208)
				Violence victimization	− 0.091 (0.208)
Peskin et al. (2014)	No	No	24 h over 24 sessions	Violence victimization	0.288 (0.123)
				Violence perpetration	0.221 (0.121)

Table 3 (continued)

Author (date)	Discussed gender roles	Included bystander training	Program length	Outcomes	Effect size (standard error)
Peskin et al. (2019)	Yes	No	5.5 h over 13 sessions	Violence perpetration	0.172 (0.151)
				Violence victimization	0.086 (0.151)
Reidy et al. (2017)	Yes	No	25 h over 25 sessions	Violence perpetration	0.025 (0.067)
				Violence victimization	− 0.059 (0.067)
Rizzo and Houck (2019)	No	No	3 h over 4 sessions	Attitudes	0.000 (0.192)
				Violence victimization	− 0.301 (0.270)
				Violence perpetration	0.011 (0.270)
Rizzo et al. (2018)	No	Missing	13 h over 6 sessions	Attitudes	0.225 (0.193)
Sanchez-Cesareo (2002)	Yes	No	4 h over 2 sessions	Knowledge	0.046 (0.220)
Sanchez-Jimenez et al. (2018)	No	Yes	7 h over 7 sessions	Violence perpetration	− 0.052 (0.069)
				Violence victimization	− 0.022 (0.065)
Sargent et al. (2017)	No	Yes	7 h over 7 sessions	Bystander behavior	0.120 (0.104)
Schramm and Gomez-Scott (2012)	No	Missing	14 h over 14 sessions	Knowledge	2.018 (0.090)
				Violence perpetration	0.251 (0.099)
Silverman (1999)	Yes	No	3 h over 3 sessions	Knowledge	0.504 (0.066)
				Attitudes	0.313 (0.061)
				Bystander behavior	0.016 (0.042)
Southgate (2016)	No	No	10 h over 10 sessions	Attitudes	− 0.049 (0.112)
Wolfe et al. (2009)	No	No	28 h over 21 sessions	Violence perpetration	0.131 (0.472)

[a] Level 2. Temporal sequence between the program and the crime or risk outcome clearly observed, or the presence of a comparison group without demonstrated comparability to the treatment group (Sherman et al. 1998, p. 4)

[b] Level 3. A comparison between two or more comparable units of analysis, one with and one without the program (Sherman et al. 1998, p. 4)

[c] Level 4. Comparison between multiple units with and without the program, controlling for other factors, or using comparison units that evidence only minor differences (Sherman et al. 1998, p. 5)

[d] Level 5. Random assignment and analysis of comparable units to program and comparison groups (Sherman et al. 1998, p. 5)

[e] Mean age of sample not provided

[f] Ethnic composition of sample not provided

[g] Program facilitator not provided

[h] Used a single-group pre-test–post-test design for the knowledge outcome and a comparison group for the attitudes and violence perpetration outcomes

[i] Used a single-group pre-test–post-test design for the knowledge outcome and a comparison group for the violence perpetration outcome

Table 4 Study and program characteristics ($n = 38$)

Characteristic	*N* (%)
Publication year (range 1997–2019)	
Pre-2015	19 (50.0)
2015–2019	19 (50.0)
Publication type	
Journal	31 (81.6)
Report	2 (5.3)
Dissertation/thesis	5 (13.2)
Randomized research design	
Randomized control trial	22 (57.9)
Quasi-experiment with matched comparison group	5 (13.2)
Single-group design	
Single-group pre-test–post-test	11 (28.9)
Two-group design	27 (71.1)
Methodological rigor	
Maryland Scale "2"	11 (28.9)
Maryland Scale "3"	2 (5.3)
Maryland Scale "4"	3 (7.9)
Maryland Scale "5"	22 (57.9)
Time of post-test	
Immediate	19 (50.0)
1–3-month post-treatment end	10 (26.3)
4–6-month post-treatment end	4 (10.5)
7–12-month post-treatment end	2 (5.3)
12 months+	3 (7.9)
Treatment group sample size at pre-test	M = 332.7 (SD = 318.2)
Treatment group sample size at post-test	M = 304.0 (SD = 281.9)
Participant age	M = 14.5 (SD = 1.2)
Race/ethnicity of analysis sample	
Caucasian/mixed	11 (28.9)
Minority	21 (55.3)
Missing	6 (15.8)
Number of hours	
< 2	6 (15.8)
2–10	23 (60.5)
11–20	4 (10.5)
> 20	5 (13.2)
Program setting	
Classroom	28 (73.7)
Small group	6 (15.8)
Other	4 (10.5)
Addressed gender roles	
No	23 (60.5)

Table 4 (continued)

Characteristic	*N* (%)
Yes	12 (31.6)
Missing	3 (7.9)
Incorporated bystander training	
No	20 (52.6)
Yes	11 (28.9)
Missing	7 (18.4)

Knowledge of dating violence ($N = 16$)

Sixteen independent effect sizes for outcomes examining the impact of a dating violence program on knowledge of dating violence were pooled (see Fig. 2). The pooled effect size of 0.566 ($z = 3.59$, $p < .001$) is a statistically significant, positive result, and suggests that dating violence prevention programs are effective at increasing adolescents' knowledge about dating violence. Significant heterogeneity was observed ($Q = 453.23$, $I^2 = 96.7\%$). No evidence of publication bias was found using Egger's test of small study effects.

Eight moderator variables were tested using subgroup analysis, and seven were significant (treatment group sample size was not a significant moderating variable) (see Table 5 for an overview of the moderator results across the four outcomes examined

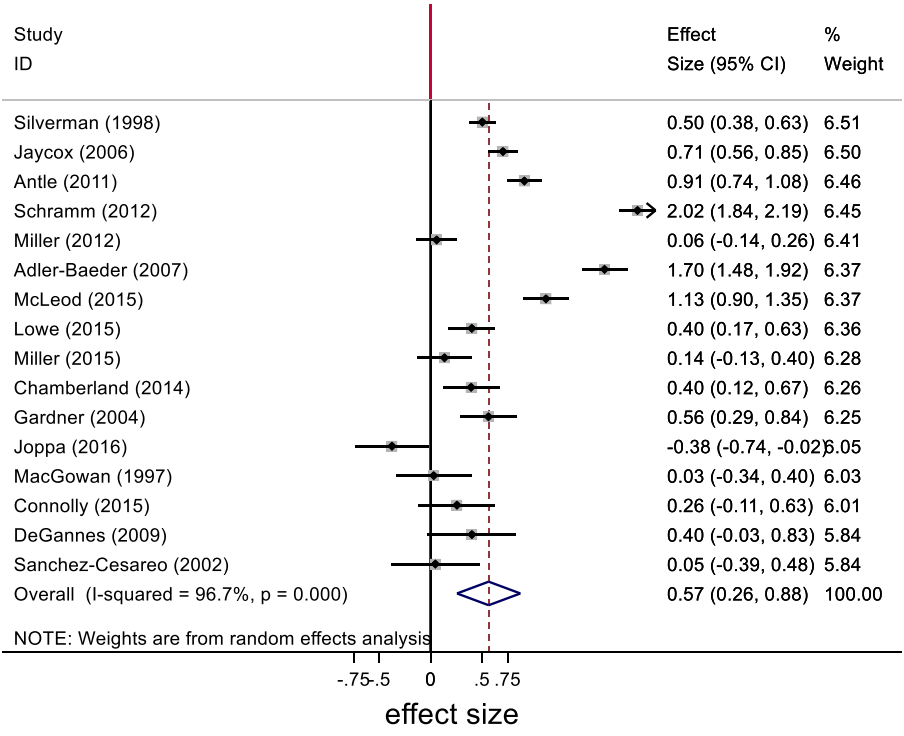


Fig. 2 Forest plot for the outcome of dating violence knowledge ($n = 16$)

Table 5 Summary of subgroup analyses by outcome

	Outcome			
	Knowledge	Attitudes	Perpetration	Victimization
Program characteristics				
Gender roles				
No	$ES = .946$, $z = 25.4$***	$ES = .104$, $z = 3.2$**	$ES = .192$, $z = 5.3$***	$ES = .135$, $z = 3.7$***
Yes	$ES = .380$, $z = 7.0$*** $Q_B = 122.2$***	$ES = .323$, $z = 6.3$*** $Q_B = 20.4$***	$ES = .055$, $z = 1.1$ $Q_B = 7.2$**	$ES = -.008$, $z = 0.2$ $Q_B = 5.8$*
Bystander training				
No	N/A	N/A	$ES = .228$, $z = 5.9$***	$ES = .115$, $z = 3.0$**
Yes			$ES = .013$, $z = 0.3$ $Q_B = 13.6$***	$ES = 0.00$, $z = 0.0$ $Q_B = 13.2$***
Program length				
< 5 h	$ES = .452$, $z = 13.4$***	$ES = .166$, $z = 5.8$***	N/A	$ES = .245$, $z = 4.6$***
> 5 h	$ES = 1.33$, $z = 26.9$*** $Q_B = 215.7$***	$ES = .143$, $z = 2.9$** $Q_B = 0.2$		$ES = .013$, $z = 0.4$ $Q_B = 13.1$***
Setting				
Middle school	N/A	$ES = .113$, $z = 1.81$	N/A	$ES = .190$, $z = 2.5$*
High school		$ES = .175$, $z = 6.07$*** $Q_B = 1.0$		$ES = .063$, $z = 2.0$* $Q_B = 2.4$
Study characteristics				
Sample size				
Less than 300	$ES = .694$, $z = 19.0$***	$ES = .205$, $z = 6.8$***	$ES = .207$, $z = 3.1$*	$ES = .124$, $z = 1.8$
300 or more	$ES = .780$, $z = 18.0$*** $Q_B = 2.3$	$ES = .069$, $z = 1.6$ $Q_B = 6.7$*	$ES = .137$, $z = 4.4$*** $Q_B = 0.9$	$ES = .074$, $z = 2.3$* $Q_B = 0.4$
Randomized design				
No	$ES = .957$, $z = 27.3$***	$ES = .209$, $z = 6.5$***	$ES = .136$, $z = 3.1$**	N/A
Yes	$ES = .342$, $z = 7.5$*** $Q_B = 113.4$***	$ES = .090$, $z = 2.3$* $Q_B = 5.6$*	$ES = .159$, $z = 4.4$*** $Q_B = 0.12$	
SGPP design				
Non-SGPP	$ES = .351$, $z = 8.2$***	$ES = .116$, $z = 3.2$***	N/A	N/A
SGPP	$ES = 1.01$, $z = 27.5$*** $Q_B = 136.0$***	$ES = .200$, $z = 5.9$*** $Q_B = 2.92$		
Publication year				
Prior to 2015	$ES = .774$, $z = 25.6$***	$ES = .134$, $z = 4.6$***	$ES = .156$, $z = 3.4$**	$ES = .047$, $z = 0.9$
2015–2019	$ES = .476$, $z = 6.5$*** $Q_B = 14.3$***	$ES = .225$, $z = 4.9$*** $Q_B = 2.8$	$ES = .145$, $z = 4.1$** $Q_B = 0.04$	$ES = .097$, $z = 2.8$** $Q_B = 0.7$
Time of post-test				
Immediate	$ES = .790$, $z = 26.6$***	$ES = .156$, $z = 5.7$***	$ES = .093$, $z = 2.1$*	$ES = .004$, $z = 0.1$
1 month+	$ES = .285$, $z = 3.5$*** $Q_B = 34.3$***	$ES = .175$, $z = 3.2$** $Q_B = 0.1$	$ES = .190$, $z = 5.2$*** $Q_B = 2.9$	$ES = .142$, $z = 3.7$*** $Q_B = 5.7$*
Sample ethnicity				
Caucasian/mixed	$ES = .963$, $z = 22.4$***	$ES = .192$, $z = 5.5$***	$ES = .328$, $z = 6.9$***	N/A
Predom minority	$ES = .574$, $z = 15.2$*** $Q_B = 51.9$***	$ES = .111$, $z = 3.1$** $Q_B = 4.7$*	$ES = .091$, $z = 2.2$* $Q_B = 25.0$***	

ES effect size, Q_B *Q*-between-group statistic, *N/A* moderator not examined

*$p < .05$, **$p < .01$, ***$p < .001$

(knowledge, attitudes, violence perpetration, and victimization)). With respect to the knowledge outcome, the moderating effect of program length was statistically significant ($Q_b = 215.67$, $p < .001$), with longer programs producing larger effect sizes than shorter programs. Programs that did *not* include the topic of gender roles or stereotypes in the curriculum produced significantly larger effect sizes than programs that did include this topic ($Q_b = 122.22$, $p < .001$). Studies with a randomized control design produced smaller effect sizes than non-randomized designs ($Q_b = 113.35$, $p < .001$), while studies using a single-group pre-test–post-test design produced significantly larger effect sizes than those with two-group designs ($Q_b = 136.0$, $p < .001$). Studies published prior to 2015 had larger effect sizes than those published more recently ($Q_b = 14.30$, $p < .001$). The timing of post-test was also a significant moderator; those studies using an immediate post-test produced larger effect sizes than those with longer follow-up times ($Q_b = 34.25$, $p < .001$). Last, studies with a primarily Caucasian or mixed ethnicity sample had larger effect sizes than those with a predominantly ethnic minority sample ($Q_b = 51.88$, $p < .001$).

Attitudes and beliefs towards dating violence ($N = 20$)

Twenty effect sizes measuring the outcome of attitudes and beliefs towards dating violence were pooled, resulting in a statistically significant, positive effect size of 0.191 ($z = 3.88$, $p < .001$) (see Fig. 3). These results suggest that youths who participate in

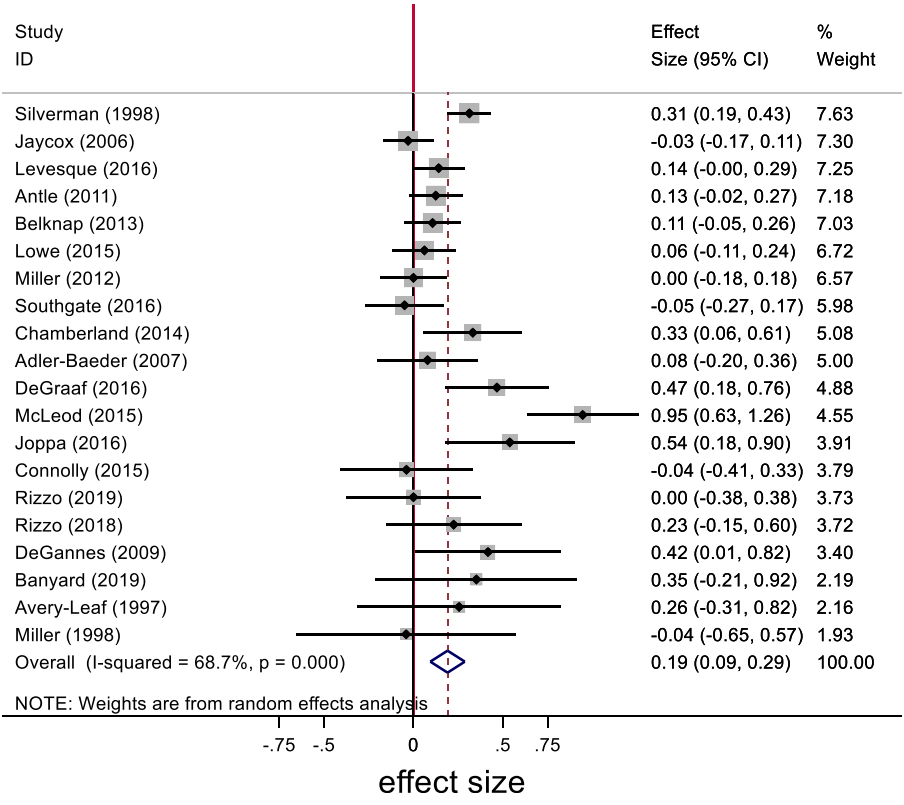


Fig. 3 Forest plot for the outcome of attitudes towards dating violence ($n = 20$)

dating violence prevention programs are less likely to be accepting of dating violence behaviors. The statistically significant Q-statistic of 60.67 and I^2 of 68.7% illustrate a considerable amount of heterogeneity within the sample. Egger's test of small study effects was not significant, indicating that publication bias is likely not a concern.

Subgroup analysis was conducted using nine moderator variables. Programs that addressed gender roles in the curriculum produced significantly larger effect sizes than those that did not ($Q_b = 20.41$, $p < .001$). Effect sizes were also larger for studies with a treatment group sample size under 300 participants versus a larger sample ($Q_b = 6.69$, $p < .05$), as well as for those with a non-randomized research design versus a randomized design ($Q_b = 5.59$, $p < .05$). Additionally, effect sizes produced by studies with predominantly Caucasian or mixed samples were significantly larger than those with predominant minority samples ($Q_b = 4.66$, $p < .05$).

Perpetration of dating violence behaviors ($N = 16$)

The pooled effect for 17 effect sizes on the perpetration of violent behaviors was significant and positive at 0.157 ($z = 3.11$, $p < .01$). These results demonstrate that incidents of dating violence among adolescents can be reduced and prevented with the implementation of dating violence prevention programs. Significant heterogeneity was again evident in the model ($Q = 42.30$, $I^2 = 62.2\%$) (see Fig. 4). The coefficient produced by Egger's test of small study effects was not statistically significant, suggesting publication bias is unlikely.

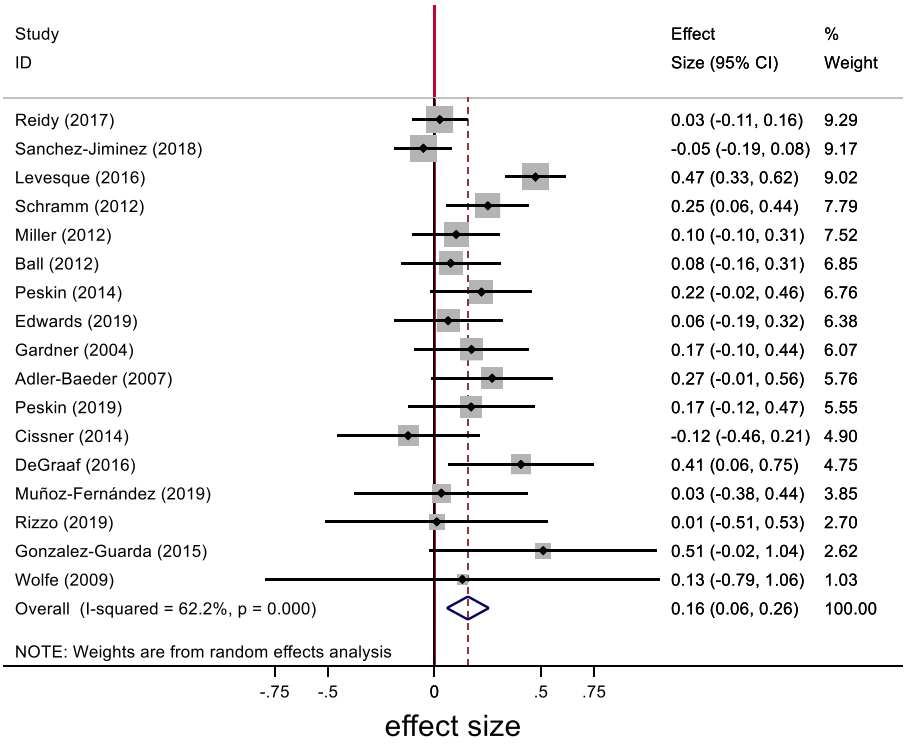


Fig. 4 Forest plot for the outcome of violence perpetration ($n = 17$)

Seven variables were investigated using subgroup analysis, three of which were significant. Significantly larger effect sizes were produced by programs that did not discuss gender roles compared to those that did ($Q_b = 7.22$, $p < .01$); similarly, programs that did not include bystander training had larger effect sizes than those that included such training ($Q_b = 13.58$, $p < .001$). Last, sample ethnicity was a significant moderator, with predominantly Caucasian or mixed samples producing larger effect sizes than minority samples ($Q_b = 24.96$, $p < .001$).

Dating violence victimization experiences ($N = 12$)

Twelve effect sizes measuring incidents of dating violence victimization yielded a non-significant pooled effect of 0.102 ($z = 1.61$, $p = 0.107$; $Q = 46.53$, $I^2 = 74.2\%$) (see Fig. 5). Unlike the previous three findings, this result suggests that dating violence prevention programs do not have a significant impact on dating violence victimization among adolescents. Egger's test of small study effects was conducted; no evidence of publication bias was found.

Four of seven variables were significant moderators in the subgroup analysis. Program length had a significant moderating effect, with shorter programs resulting in larger effect sizes ($Q_b = 13.12$, $p < .001$). Programs also produced larger, positive effect sizes if they did not discuss gender roles ($Q_b = 5.82$, $p < .05$), or include bystander training ($Q_b = 13.19$, $p < .001$). Last, timing of the post-test was a significant moderator, with larger effect sizes for longer follow-ups compared to immediate post-tests ($Q_b = 5.65$, $p < .05$).

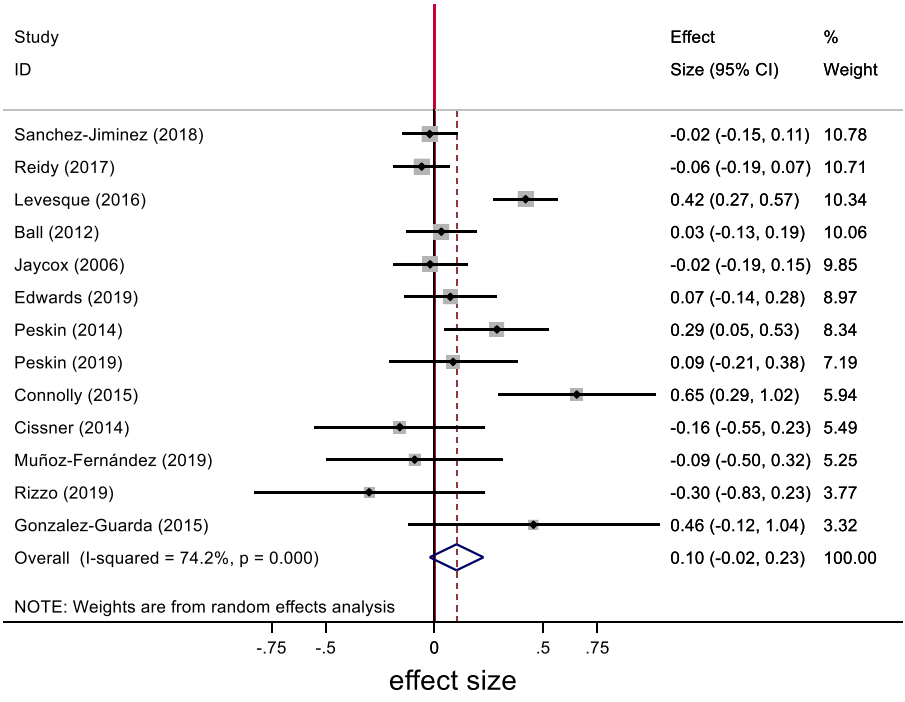


Fig. 5 Forest plot for the outcome of violence victimization ($n = 13$)

Bystander behaviors and intentions ($N = 6$)

Seven effect sizes focused on bystander behaviors and intentions were identified; however, the influence analysis identified one outlier study (Plourde et al. 2015). This outlier was removed and the remaining six effect sizes were pooled, resulting in a non-significant estimate of 0.120 ($z = 1.90$, $p = .058$) (see Fig. 6). This result implies that dating violence prevention programs for adolescents are only marginally effective at increasing bystander behaviors and intentions. Given the non-significant Q-statistic of 9.33, along with the small sample size for this outcome measure, moderator analyses were not performed. Egger's test of small study effects provides no indication of bias.

Discussion

The current study examined the overall effectiveness of adolescent dating violence prevention programs at increasing knowledge about dating violence, improving attitudes towards dating violence, reducing incidents of dating violence perpetration and victimization, and increasing the prevalence of bystander behaviors and intentions. As a whole, programs appear effective at increasing adolescents' knowledge about dating violence behaviors and impacts, as well as changing attitudes and beliefs concerning dating violence. These findings are consistent with prior meta-analyses, despite minimal overlap of included studies across the meta-analytic samples. Although the effect sizes are somewhat small, most prevention programs seem to have a positive impact on adolescents when it comes to increasing knowledge of and changing attitudes towards dating violence. Several treatment and study characteristics were found to be significant

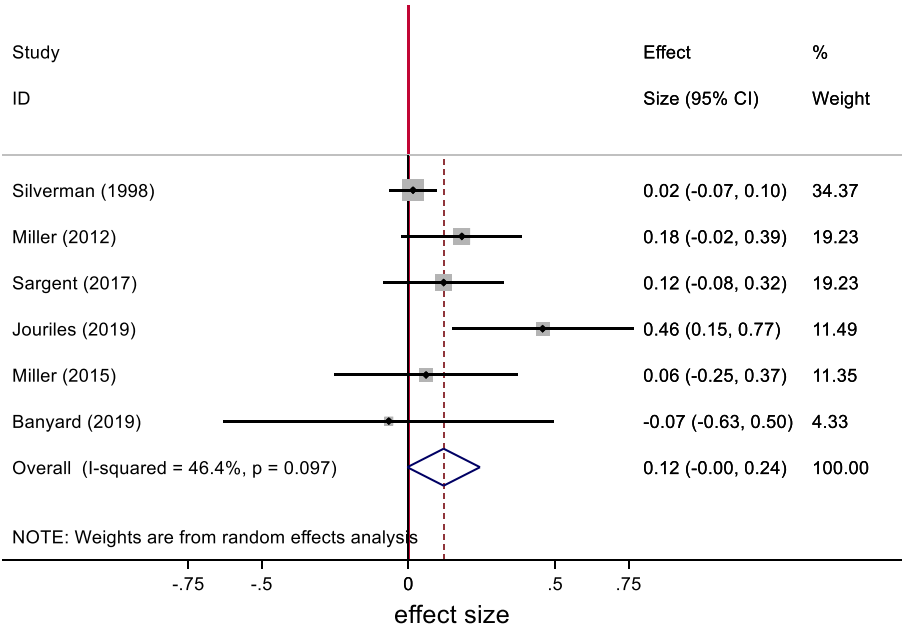


Fig. 6 Forest plot for the outcome of bystander behavior ($n = 6$)

moderators of these outcomes, including the discussion of gender roles, timing of post-test, and research design type. When examining attitudes towards dating violence, programs that included a focus on gender roles produced larger effect sizes than programs that did not. This finding is expected; given that gender roles and stereotypes are believed to perpetuate myths regarding gender-based violence (Sampert 2010), we would anticipate that teaching adolescents about these topics could help to dispel myths and shift attitudes in a more positive direction. Contrary to expectations, the discussion of gender roles produced the opposite effect for the knowledge outcome; programs that included this topic produced significantly smaller effect sizes than did programs which omitted it. While we can only postulate as to the reason for this unanticipated finding, it may be due to discrepancies between the program content and outcome measures. If a program includes the topic of gender roles and stereotypes, less time might be spent on general dating violence knowledge. However, if the knowledge outcome measures focus on more general violence topics (e.g., warning signs of abuse; Silverman 1999), they may not be suitably tailored to the content presented, resulting in smaller effect sizes.

Regarding study characteristics, studies with an immediate post-test produced significantly larger effects for the knowledge outcome than those with a longer time to follow-up. This result is consistent with expectations, as program effects tend to be strongest immediately after a program ends and dissipate as time progresses (de la Rue et al. 2017). Additionally, studies with randomized designs produced significantly smaller effect sizes for the knowledge and attitude outcomes than did studies with non-randomized designs. Similarly, studies with single-group pre-test–post-test designs produced significantly larger effect sizes. These findings are not unexpected as it is well documented that less rigorous research designs are more likely to overestimate the treatment effect (Card 2011; Lipsey and Wilson 2001; Morris and DeShon 2002). Nevertheless, this result is important given the high number of studies that employed a single-group pre-test–post-test design. While significant differences exist between the effect sizes for randomized vs. non-randomized designs and single-group vs. two-group designs, all designs produced a positive, significant effect suggesting that the treatment impact is sufficiently robust to be detected by both moderate and rigorous design types.

While knowledge and attitude outcomes generally had positive effects, a different pattern of results emerged with respect to behaviors. When examining incidents of dating violence perpetration (of any psychological/emotional, physical, or sexual violence), a positive treatment impact was found with participants exhibiting a significant decrease in perpetration behaviors. This result is encouraging and consistent with the aforementioned positive and significant findings on knowledge and attitudes. It is also consistent with past research demonstrating that acceptance of dating violence and more equitable gender attitudes are significant predictors of dating violence perpetration among adolescents, suggesting that changing these beliefs is an important step to changing perpetration behaviors (Foshee et al. 2001; Miller et al. 2020; Mumford et al. 2020; McNaughton Reyes et al. 2016; Taylor et al. 2015). Conversely, the analyses of program impact on both dating violence victimization and bystander behavior/intentions resulted in positive but non-significant pooled effects. These results are somewhat unexpected; similar to the perpetration of dating violence, previous research suggests that acceptance of dating violence is a significant predictor of dating violence victimization (Fernández-Antelo et al. 2020; Herrenkohl and Jung 2016; Karlsson et al.

2016; Machado et al. 2010; Simon et al. 2010). Thus, if anti-dating violence attitudes are increasing, we would expect victimization rates to decrease as adolescents become less likely to tolerate this type of behavior. It may be that dating violence programs can successfully teach students that violence in relationships is unacceptable, yet challenges remain for adolescents in their ability to change the abusive behaviors of their partners. While the victimization outcome was non-significant overall, several significant moderator variables were identified—warranting a closer examination of these studies to determine whether any important programmatic differences are seen.

The inclusion of bystander training and the topic of gender roles were both significant moderators for the victimization outcome. Programs that addressed gender roles produced significantly different and negative effect sizes compared to those that did not address gender roles. This negative trend may be due to adolescents developing an increased ability to recognize multiple forms of dating violence victimization. For example, participants may only report physical violence at pre-test; however, the discussion of gender stereotypes and gender-related psychological abuse may lead participants to identify additional incidents of violence that they might not have considered at pre-test. As such, reports of victimization are increased at post-test, resulting in what appears to be a negative program effect.

The inclusion of discussion surrounding gender roles and inequities also produced significantly smaller effect sizes for the perpetration outcome. When these results are taken together in the context of the victimization outcome, it implies that perhaps the delivery of this particular information is not effective at impacting behaviors. This type of content can be sensitive, with the risk of being portrayed as critical or placing blame; research in sexual violence prevention has found that programs can lead participants to feel as though they are being told how to think and act, and thus, they react with hostility (Malamuth et al. 2018; Spikes and Sternadori 2018). The presentation of the gender role and equity content in dating violence programs may elicit similar defensive or negative reactions from participants, resulting in smaller or negative effects for both perpetration and victimization. Adolescents may also have difficulty in translating what they learn about gender roles and inequities into changes in their behaviors. If the teaching methods do not actively demonstrate how to incorporate the information into their behaviors and choices, participants may not see the direct connections between these topics and their actions.

Additionally, programs that included bystander training also produced null effects for the victimization outcome; this result is again counter to expectation. One potential explanation may be that bystander training does not impact participants' own experiences of victimization. In other words, bystander training is focused on intervening in situations involving other people and diffusing these situations (Coker et al. 2017; Cook-Craig et al. 2014); the same techniques may not necessarily apply when the individual is directly involved in a situation as the target of the abuse.

The effects on victimization were also moderated by program length, with the unexpected finding that shorter programs produce significantly larger effects, though shorter programs have been found to be associated with larger effects sizes in other areas of violence prevention (DeVries et al. 2015; van der Put et al. 2018). It may be that shorter programs are more direct and succinct with respect to educating participants on how to recognize and prevent personal victimization. Perhaps participants are more likely to pay attention to the material when it is presented in a shorter time frame

(Flores and Hartlaub 1998). This is particularly likely with younger adolescents who may be more prone to boredom or fatigue as the program progresses. Evidence also suggests that the dosage level of a program should be matched with the risk level and needs of the target audience (Nation et al. 2003). The dosage/participant risk level of these programs may be mismatched; for example, some programs may have targeted a low-risk audience that does not require a lengthy program.

It is also worthwhile to consider the noteworthy finding that the pooled treatment effect on victimization was null, while the pooled effect on violence perpetration was significant and positive. We might expect that these outcomes would mirror one other; if perpetration is significantly decreasing, likewise should victimization decrease (i.e., if adolescents are perpetrating less violence, there fewer adolescents should be victimized). However, this does not appear to be the case. This finding may be due to reporting practices; perpetration of dating violence is socially undesirable, a perspective which is likely to be even more evident following participation in a prevention program. It may be that participants are less willing to report their continued perpetration of these behaviors after completing the program, yet the likelihood of participants reporting victimization remains consistent or increases (Fernández-González et al. 2013; Visschers et al. 2017). Alternatively, participants may have been underreporting their victimization experiences at the time of pre-test if they did not realize their experiences constituted dating violence. If the programs are successful at increasing knowledge about dating violence, participants may become able to identify instances of violence at post-test that they did not classify as such at pre-test, thus inflating the post-test measure and masking any possible decreases. This finding highlights the importance of considering both perpetration and victimization as separate outcomes, with one not necessarily a direct reflection of the other.

Bystander behavior was also not significantly impacted by dating violence prevention programs, though the direction of the effect was generally positive. It is also noteworthy that sensitivity testing on the bystander behavior outcome indicated that when the one study using a single-group pre-test–post-test design is removed from the pooled analysis, the remaining pooled effect (from five two-group design studies) was positive and significant. This result implies that the single-group pre-test–post-test study negatively impacts the pooled effect size, which is contrary to the expectation that single-group designs are weaker and bias effect sizes upwards. However, the bystander behavior outcome sample is small with a small pooled effect size at best; as such, evidence suggests the overall effect of dating violence programs on bystander outcomes is null to weak overall.

The mixed results regarding the behavioral impacts of dating violence prevention programs suggest that behaviors are more difficult to change than knowledge or beliefs, which is consistent with developmental and behavioral research (Kelly and Barker 2016; McMaster et al. 2002). However, it is also important to consider the time of follow-up for these particular outcomes—most of the effect sizes were based on measures assessed at an immediate post-test. Knowledge and attitudes may reflect an immediate change in program participants, as they do not necessarily require any implementation or practice. Behavioral outcomes, on the other hand, may take time to evidence any noticeable change (Coker et al. 2017; Kelly and Barker 2016). This is in part due to the lack of opportunity for participants to implement newly acquired knowledge in situations in which these behaviors may occur. At the time of an

immediate post-test, a participant may not yet have experienced any opportunity to engage in bystander behaviors.

Limitations

The first limitation of the current research is a lack of reported data in the primary evaluation studies, precluding a detailed assessment of all relevant program components that might be associated with program success or failure. For example, all variables relating to program components (such as whether the program used group discussion or included skills rehearsal) were missing some level of data due to vagueness in the studies' descriptions of the interventions. As such, many moderators were unusable in the subgroup analyses or were used but with small sample sizes per subgroup (resulting in a potential lack of power to detect impact). In fact, we initially attempted to code for a great deal more treatment-level characteristics in the current research, but it became evident during the coding phase that very few characteristics were reported on consistently by primary study authors. Given that the effectiveness of a dating violence prevention program may be largely dependent on the specific curriculum used, our inability to fully consider all of these characteristics is unfortunate. We note that this concern is not limited to the current study, however, but is often a problem for meta-analysis more generally due to evaluation reports differing wildly in the level of detail they present about the programs themselves. In other words, meta-analysts are left to rely on authors' often short descriptions of program content, which may fail to present key information on treatment characteristics.

Additionally, the small sample sizes used in the subgroup analyses suggest caution is warranted when considering the policy and practice implications of the results. With small subgroup sample sizes, the likelihood of any individual study exerting a substantial influence on the pooled effect size increases; individual studies in the pooled fixed effects models are weighted by their inverse standard error which is a function of sample size—by definition, this means larger studies contribute more to the pooled effect. As such, it is important to interpret these results and conclusions regarding the moderator variables with care. Despite their potential volatility, these analyses are important to consider; program and study characteristics are important factors when considering the effectiveness of these programs, and the insight afforded by these analyses is beneficial for our overall understanding of evidence-based practices for dating violence prevention programs.

An additional potential limitation of the selection criteria is the decision rule to prioritize immediate post-tests rather than longer follow-up measures. Arguably, a delayed follow-up would offer a more conservative estimate of treatment effect, and the significant findings for the knowledge and attitude outcomes may have been diminished. Conversely, as discussed, behavioral impacts may be less detectable at an immediate post-test. Due to differences across study designs, prioritizing the immediate post-test maximized commensurability between effect sizes. The delayed effect is important to investigate and should be examined in future research.

There is also a possible limitation regarding the choice of ICC estimates for the cluster adjustments. The literature on specific ICC guidelines for attitudinal and behavioral outcomes in an adolescent population is severely lacking. We searched the existing literature for empirically reported ICCs of relevant outcomes; by using a range of prior estimates to develop our ICCs, we believe those applied in the current study are well-informed. Additionally, many existing meta-analyses fail to adjust for clustering altogether, which can lead to misestimation of the treatment effect. Including cluster adjustments using the best available estimates ensures that within-study nesting is accounted for.

Arguably, the decision to include single-group, pre-test–post-test designs is a limitation of the current meta-analysis. There is mixed support in the literature for whether or not it is appropriate to combine effect sizes from single-group and two-group designs (see Borenstein and Hedges 2019; Carlson and Schmidt 1999; Cuijpers et al. 2017; Morris and DeShon 2002). Our decision was based on the work of Morris and DeShon (2002), who argue that single-group and two-group designs can be pooled after several methodological considerations. The inclusion of both types of designs in the present study differentiates this meta-analysis from previous research and we believe it allows for a more comprehensive examination of the literature.

Conclusions

Overall, dating violence prevention programs for adolescents appear to be effective at improving knowledge and attitudinal measures of dating violence, but are not as successful at improving behaviors. Given that existing research has found knowledge and attitudes to be significant predictors of dating violence behaviors (e.g., Karlsson et al. 2016; McNaughton Reyes et al. 2016), these findings suggest that successful behavioral changes are possible but may not be evidenced in short-term measures of behavioral outcomes. Future research should further investigate the relationship between short-term improvements in knowledge and attitudes with respect to dating violence and long-term behavioral changes. While results provide support for the continued implementation of dating violence programs with an adolescent population, they also suggest that programs are not created equal and program characteristics (e.g., content covered, intervention length) may influence the levels of impact seen. Intimate partner violence is a problem that often begins in adolescence, and it is imperative that programs designed to address these behaviors early and prevent their persistence are in fact achieving this goal. Given the lack of consistent reporting, future evaluation research should also ensure that comprehensive reporting methods are employed to allow for a detailed understanding of the content of dating violence prevention programs. This information would enable the examination of a greater number of moderators of treatment impact, resulting in recommendations concerning what type of program content is related to stronger or weaker program impacts. An increased understanding of effective dating violence prevention efforts will aid policymakers and program developers in creating successful primary prevention strategies and resources for educators.

Appendix. List of electronic databases searched

- Academic Search Premier
- BioMed Central
- Canadian Research Index
- CINAHL Complete
- Cochrane Central Register of Controlled Trials
- Cochrane Database of Systematic Reviews
- Criminal Justice Abstracts with Full Text
- Database of Abstracts of Reviews of Effects
- Education Source
- Government of Canada Publications
- MEDLINE (OVID)
- MEDLINE with Full Text
- National Criminal Justice Reference Service (NCJRS)
- Open Access Theses and Dissertations
- ProQuest Dissertations and Theses Abstracts and Index
- PsycARTICLES
- PsycBOOKS
- PsycINFO
- Social Sciences Abstracts (H.W. Wilson)
- Social Sciences Full Text (H.W. Wilson)
- Sociological Collection
- Web of Science
- Women's Studies International

References

Asterisk denotes studies that are included in the meta-analysis.

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

*Adler-Baeder, F., Krepelman, J. L., Schramm, D. G., Higginbotham, B., & Paulk, A. (2007). The impact of relationship education on adolescents of diverse backgrounds. *Family Relations, 56*, 291–303.

*Antle, B. F., Sullivan, D. J., Dryden, A., Karam, E. A., & Barbee, A. P. (2011). Healthy relationship education for dating violence prevention among high-risk youth. Children and Youth Services Review, 33(1), 173–179. DOI: https://doi.org/10.1016/j.childyouth.2010.08.031.

*Avery-Leaf, S., Cascardi, M., O'Leary, K. D., & Cano, A. (1997). Efficacy of a dating violence prevention program on attitudes justifying aggression. Journal of Adolescent Health, 21,11–17.

*Ball, B., Tharp, A. T., Noonan, R. K., Valle, L. A., Hamburger, M. E., & Rosenbluth, B. (2012). Expect respect support groups: preliminary evaluation of a dating violence prevention program for at-risk youth. Violence Against Women, 18(7), 746–762. DOI: https://doi.org/10.1177/1077801212455188.

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

*Banyard, V. L., Edwards, K. M., Rizzo, A. J., Theodores, M., Tardiff, R., Lee, K., & Greenberg, P. (2019). Evaluating a gender transformative violence prevention program for middle school boys: a pilot study.

Children and Youth Services Review, 101, 165–173. DOI: https://doi.org/10.1016/j.childyouth.2019.03.052.

*Belknap, R. A., Haglund, K., Felzer, H., Pruszynski, J., & Schneider, J. (2013). A theater intervention to prevent teen dating violence for Mexican-American middle school students. Journal of Adolescent Health, 53(1), 62–67. DOI: https://doi.org/10.1016/j/jadohealth.2013.02.006.

Black, B. M., Hawley, A., Hoefer, R., & Barnett, T. M. (2017). Factors related to teenage dating violence prevention programming in schools. *Children & Schools, 39*(2), 99–107. https://doi.org/10.1093/cs/cdx007.

Borenstein, M., & Hedges, L. V. (2019). Statistically describing and combining study outcomes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 207–243). New York: Russell Sage Foundation.

Brackett, M. A., Reyes, M. R., Rivers, S. E., Elbertson, N. A., & Salovey, P. (2011). Classroom emotional climate, teacher affiliation, and student conduct. *Journal of Classroom Interaction, 46*(1), 27–36.

Brown, A., Cosgrave, E., Killackey, E., Purcell, R., Buckby, J., & Yung, A. R. (2009). The longitudinal association of adolescent dating violence with psychiatric disorders and functioning. *Journal of Interpersonal Violence, 24*(12), 1964–1979. https://doi.org/10.1177/0886260508327700.

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

Card, N. A. (2011). *Applied meta-analysis for social science research*. New York: The Guilford Press.

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: findings from the research literature on training. *Journal of Applied Psychology, 84*(6), 851–862.

Cascardi, M., King, C. M., Rector, D., & DelPozzo, J. (2018). School-based bullying and teen dating violence prevention laws: overlapping or distinct? *Journal of Interpersonal Violence, 33*(21), 3267–3297. https://doi.org/10.1177/0886260518798357.

Centers for Disease Control and Prevention (2019) Preventing teen dating violence. Retrieved from: https://www.cdc.gov/violenceprevention/intimatepartnerviolence/teendatingviolence/fastfact.html

*Chamberland, A., Cantin-Drouin, M., & Damant, D. (2014). Assessment of the impact of SAISIR: a dating violence prevention program. *The Canadian Social Work Review, 31*(1),125–139.

*Cissner, A. B., & Hassoun Ayoub, L. (2014). *Building healthy teen relationships: an evaluation of the fourth R curriculum with middle school students in the Bronx*. Report submitted to the National Institute of Justice. Retrieved from: https://www.ncjrs.gov/pdffiles1/nij/grants/248486.pdf

Coker, A. L., McKeown, R. E., Sanderson, M., Davis, K. E., Valois, R. F., & Huebner, E. S. (2000). Severe dating violence and quality of life among South Carolina high school students. *American Journal of Preventive Medicine, 19*(4), 220–228.

Coker, A. L., Bush, H., Cook-Craig, P., DeGue, S. A., Clear, E. R., Brancato, C. J., Fisher, B. S., & Recktenwald, E. A. (2017). RCT testing bystander effectiveness to reduce violence. *American Journal of Preventative Medicine, 52*(5), 566–578.

Community Preventative Services Task Force (2018). *Violence: primary prevention interventions to reduce perpetration of intimate partner violence and sexual violence among youth.* Retrieved from: https://www.thecommunityguide.org/findings/violence-primary-prevention-interventions-reduce-perpetration-intimate-partner-violence-sexual-violence-among-youth

*Connolly, J., Josephson, W., Schnoll, J., Simkins-Strong, E., Pepler, D., MacPherson, A., Weiser, J., Moran, M., & Jiang, D. (2015). Evaluation of a youth-led program for preventing bullying, sexual harassment, and dating aggression in middle schools. *The Journal of Early Adolescence, 35*(3), 403–434. https://doi.org/10.1177/0272431614535090.

Cook-Craig, P. G., Coker, A. L., Clear, E. R., Garcia, L. S., Bush, H. M., Brancato, C. J., Williams, C. M., & Fisher, B. S. (2014). Challenge and opportunity in evaluating a diffusion-based active bystanding prevention program: green dot in high schools. *Violence Against Women, 20*(10), 1179–1202.

Crooks, C. V., Jaffe, P., Dunlop, C., Kerry, A., & Exner-Cortens, D. (2019). Preventing gender-based violence among adolescents and young adults: lessons from 25 years of program development and evaluation. *Violence Against Women, 25*(1), 29–55. https://doi.org/10.1177/1077801218815778.

Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiological Psychiatric Science, 26*(4), 364–368.

*de Graaf, I., de Haas, S., Zaagsma, M., & Wijsen, C. (2016). Effects of rock and water: an intervention to prevent sexual aggression. *Journal of Sexual Aggression, 22*(1), 4–19. https://doi.org/10.1080/13552600.2015.1023375.

De Koker, P., Matthews, C., Zuch, M., Bastien, S., & Mason-Jones, A. J. (2014). A systematic review of interventions for preventing adolescent intimate partner violence. *Journal of Adolescent Health, 54*, 3–13. https://doi.org/10.1016/j.jadohealth.2013.08.008.

de la Rue, L., Polanin, J. R., Espelage, D. L., & Pigott, T. D. (2017). A meta-analysis of school-based interventions aimed to prevent or reduce violence in teen dating relationships. *Review of Educational Research, 87*(1), 7–34. https://doi.org/10.3102/0034654316632061.

*DeGannes, A. B. (2009). *Evaluation of sexual assault and dating violence prevention program for middle school students*. PhD dissertation, Temple University.

DeGue, S., Valle, L. A., Holt, M. K., Massetti, G. M., Matjasko, J. L., & Teten Tharp, A. (2014). A systematic review of primary prevention strategies for sexual violence perpetration. *Aggression and Violent Behavior, 19*, 346–362. https://doi.org/10.1016/j.avb.2014.05.004.

DeVries, S. L. A., Hoeve, M., Assink, M., Stams, G. J. J. M., & Asscher, J. (2015). Practitioner review effective ingredients of prevention programs for youth at risk of persistent juvenile delinquency – recommendations for clinical practice. *The Journal of Child Psychology and Psychiatry, 52*(2), 108–121. https://doi.org/10.1111/jcpp.12320.

Di Stasio, M. R., Savage, R., & Burgos, G. (2016). Social comparison, competition and teacher-student relationships in junior high school classrooms predicts bullying and victimization. *Journal of Adolescence, 53*, 207–216. https://doi.org/10.1016/j.adolescence.2016.10.002.

Edwards, S. R., & Hinsz, V. B. (2014). A meta-analysis of empirically tested school-based dating violence prevention programs. *SAGE Open, 4*(2), 1–8. https://doi.org/10.1177/21582444014535787.

*Edwards, K. M., Banyard, V. L., Sessarego, S. N., Waterman, E. A., Mitchell, K. J., & Chang, H. (2019). Evaluation of a bystander-focused interpersonal violence prevention program with high school students. Prevention Science, 20(4), 488–498. DOI: https://doi.org/10.1007/s11121-019-01000-w.

Egger, M., & Smith, G. D. (2001). Principles and procedures for systematic reviews. In M. Egger, G. Smith, & D. G. Altman (Eds.), *Systematic reviews in health care: meta-analysis in context* (2nd ed., pp. 23–42). London: BMJ Publishing Group.

Fellmeth, G., Heffernan, C., Nurse, J., Habibula, S., & Sethi, D. (2015). Educational and skills-based interventions to prevent relationship violence in young people. *Research on Social Work Practice, 25*(1), 90–102 https://doi.org/10.1177/1049731514533392.

Fernández-Antelo, I., Cuadrado-Gordillo, I., & Martin Mora Parra, G. (2020). Synergy between acceptance of violence and sexist attitudes as a dating violence risk factor. *International Journal of Environmental Research and Public Health, 17*, 1–12. https://doi.org/10.3390/ijerph17145209.

Fernández-González, L., O'Leary, K. D., & Muñoz-Rivas, M. J. (2013). We are not joking: need for controls in reports of dating violence. *Journal of Interpersonal Violence, 28*(3), 602–620. https://doi.org/10.1177/0886260512455518.

Flores, S., & Hartlaub, M. G. (1998). Reducing rape-myth acceptance in male college students: a meta-analysis of intervention studies. *Journal of College Student Development, 39*(5), 438–448.

Fonagy, P., Twemlow, S. W., Vernberg, E. M., Nelson, J. M., Dill, E. J., Little, T. D., & Sargent, J. A. (2009). A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression. *Journal of Child Psychology and Psychiatry, 50*(5), 607–616. https://doi.org/10.1111/j.1469-7610.2008.02025.x

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

Foshee, V. A., McNaughton Reyes, H. L., Gottfredson, N. C., Chang, L., & Ennett, S. T. (2013). A longitudinal examination of psychological, behavioral, academic, and relationship consequences of dating abuse victimization among a primarily rural sample of adolescents. *Journal of Adolescent Health, 53*(6), 723–729. https://doi.org/10.1016/j.jadohealth.2013.06.016.

*Gardner, S. P., Giese, K., & Parrott, S. M. (2004). Evaluation of the connections: relationships and marriage curriculum. Family Relations, 53(5), 521–527.

*Gonzalez-Guarda, R. M., Guerra, J. E., Cummings, A. A., Pino, K., & Becerra, M. M. (2015). Examining the preliminary efficacy of a dating violence prevention program for Hispanic adolescents. Journal of School Nursing, 31(6), 411–421. DOI: https://doi.org/10.1177/1059840515598843.

Graham, L. M., Embry, V., Young, B., Macy, R. J., Moraccco, K. E., McNaughton Reyes, H. L., & Martin, S. L. (2019). Evaluations of prevention programs for sexual, dating, and intimate partner violence for boys and men: a systematic review. *Trauma, Violence & Abuse*, 1–27. https://doi.org/10.1177/1524838019851158.

Hedges, L. V. (2000). Meta-Analysis. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 5, pp. 202–204). New York: Oxford University Press.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60–87.

Herrenkohl, T. I., & Jung, H. (2016). Effects of child abuse, adolescent violence, peer approval and pro-violence attitudes on intimate partner violence in adulthood. *Criminal Behaviour and Mental Health, 26*, 304–314. https://doi.org/10.1002/cbm.2014.

Hickman, L. J., Jaycox, L. H., & Aronoff, J. (2004). Dating violence among adolescents: prevalence, gender distribution, and prevention program effectiveness. *Trauma Violence Abuse, 5*(2), 123–142. https://doi.org/10.1177/1524838003262332.

Hoefer, R., Black, B., & Richard, M. (2015). The impact of state policy on teen dating violence prevalence. *Journal of Adolescence, 44*, 88–96. https://doi.org/10.1016/j.adolescence.2015.07.006.

Hotton Mahoney, T. (2010). *Police reported dating violence in Canada, 2008*. Juristat. Statistics Canada. Retrieved from: https://www150.statcan.gc.ca/n1/pub/85-002-x/2010002/article/11242-eng.htm.

Isaacs, J., Voeten, M., & Salmivalli, C. (2013). Gender-specific or common classroom norms? Examining the contextual moderators of the risk for victimization. *Social Development, 22*(3), 555–579. https://doi.org/10.1111/j.1467-9507.2012.00655.x.

*Jaycox, L. H., McCaffrey, D., Eiseman, B., Aronoff, J., Shelley, G. A., Collins, R. L., & Marshall, G. A. (2006). Impact of a school-based dating violence prevention program among Latino teens: randomized controlled effectiveness trial. Journal of Adolescent Health, 39(5), 694–704. DOI: https://doi.org/10.1016/j.jadohealth.2006.05.002.

*Joppa, M. C., Rizzo, C. J., Nieves, A. V., Brown, L. K. (2016). Pilot investigation of the Katie Brown Educational Program: a school-community partnership. Journal of School Health, 86(4), 288–297.

*Jouriles, E. N., Sarget, K. S., Salis, K. L., Caiozzo, C., Rosenfield, D., Cascardi, M., Grych, J. H., O'Leary, K. D., & McDonald, R. (2019). Increasing bystander behavior to prevent adolescent relationship violence: a randomized controlled trial. Journal of Consulting and Clinical Psychology, 87(1), 3–15. DOI: https://doi.org/10.1037/ccp0000355.

Kann, L., McManus, T., Harris, W. A., Shanklin, S. L., Flint, K. H., Queen, B., Lowry, R., Chyen, D., Whittle, L., Thornton, J., Lim, C., Bradford, D., Yamakawa, Y., Leon, M., Brener, N., & Ethier, K. A. (2018). Youth risk behavior surveillance – United States, 2017. *Morbidity and Mortality Weekly Report. Surveillance Summaries, 67*(8), 1–113.

Karlsson, M. E., Temple, J. R., Weston, R., & Le, V. D. (2016). Witnessing interpersonal violence and acceptance of dating violence as predictors for teen dating violence. *Violence Against Women, 22*(5), 625–646. https://doi.org/10.1177/1077801215605920.

Kelly, M. P., & Barker, M. (2016). Why is changing health-related behaviour so difficult? *Public Health, 136*, 109–116.

Kerig, P. K. (2010). Adolescent dating violence in context: introduction and overview. *Journal of Aggression, Maltreatment & Trauma, 19*(5), 465–468.

Lapierre, A., Paradis, A., Todorov, E., Blais, M., & Hebert, M. (2019). Trajectories of psychological dating violence perpetration in adolescence. *Child Abuse and Neglect, 97*, 1–13. https://doi.org/10.1016/j.chiabu.2019.104167.

Leen, E., Sorbring, E., Mawer, M., Holdsworth, E., Helsing, B., & Bowen, E. (2013). Prevalence, dynamic risk factors and the efficacy of primary interventions for adolescent dating violence: an international review. *Aggression and Violent Behavior, 18*(1), 159–174. https://doi.org/10.1016/j.avb.2012.11.015.

*Levesque, D. A., Johnson, J. L., Welch, C. A., Prochaska, J. M., & Paiva, A. L. (2016). Teen dating violence prevention: cluster-randomized trial of teen choices, an online, stage-based program for healthy, nonviolent relationships. Psychology of Violence, 6(3), 421–432. DOI: https://doi.org/10.1037/vio0000049.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis. Applied social research methods series: Vol 49*. Thousand Oaks: Sage Publications.

*Lowe, L. A., Jones, C. D., & Banks, L. (2008). Preventing dating violence in public schools. Journal of School Violence, 6(3), 69–87. DOI: https://doi.org/10.1300/J202v06n03.

*MacGowan, M. J. (1997). An evaluation of a dating violence prevention program for middle school students. Violence and Victims, 12(3), 233–235.

Machado, C., Cariade, S., & Martins, C. (2010). Violence in juvenile dating relationships: self-reported prevalence and attitudes in a Portuguese sample. *Journal of Family Violence, 25*, 43–52. https://doi.org/10.1007/s10896-009-9268-x.

Malamuth, N. M., Huppin, M., & Linz, D. (2018). Sexual assault interventions may be doing more harm than good with high-risk males. *Aggression and Violent Behavior, 41*, 20–24. https://doi.org/10.1016/j.avb.2018.05.010.

*McLeod, D. A., Jones, R., & Cramer, E. P. (2015). An evaluation of a school-based, peer-facilitated, healthy relationship program for at-risk adolescents. *Children & Schools, 37*(2), 108–116.

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

McNaughton Reyes, H. L., Foshee, V. A., Niolon, P. H., Reidy, D. E., & Hall, J. E. (2016). Gender role attitudes and male adolescent dating violence perpetration: normative beliefs as moderators. *Journal of Youth and Adolescence, 45*, 350–360. https://doi.org/10.1007/s10964-015-0278-0.

*Miller, J. R. (1998). A social learning perspective toward the prevention of dating violence: an evaluation of a group counseling model. PhD dissertation. *Miami Institute of Psychology of the Caribbean Center for Advanced Studies.*

*Miller, E., Tancredi, D. J., McCauley, H. L., Decker, M. R., Virata, M. C. D., Anderson, H. A., Stetkevich, N., Brown, E. W., Moideen, F., & Silverman, J. G. (2012). 'Coaching boys into men': a cluster-randomized controlled trial of a dating violence prevention program. Journal of Adolescent Health, 51(5), 431–438. DOI: https://doi.org/10.1016/j.jadohealth.2012.01.018.

*Miller, E., Goldstein, S., McCauley, H. L., Jones, K. A., Dick, R. N., Jetton, J., Silverman, J. G., Blackburn, S., Monasterio, E., James, L., & Tancredi, D. J. (2015). A school health center intervention for abusive adolescent relationships: a cluster RCT. Pediatrics, 135(1), 76–85.

Miller, E., Jones, K. A., & Ripper, L., Paglisott, T., Mulbah, P., Abebe, K. Z. (2020). An athletic coach-delivered middle school gender violence prevention program. *JAMA Pediatrics, 174*(3), 241–240.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105–125. https://doi.org/10.1037//1082-989X.7.1.105.

Mumford, E. A., Taylor, B. G., & Giordano, P. C. (2020). Perpetration of adolescent dating relationship abuse: the role of conditional tolerance for violence and friendship factors. *Journal of Interpersonal Violence, 35*(5–6), 1206–1228. https://doi.org/10.1177/0886260517693002.

*Munoz-Fernandez, N., Ortega-Rivera, J., Nocentini, A., Menesini, E., & Sanchez-Jimenez, V. (2019). The efficacy of the "Dat-e Adolescence" prevention program in the reduction of dating violence and bullying. International Journal of Environmental Research and Public Health, 16(408), 1–21. DOI: https://doi.org/10.3390/ijerph16030408.

Murray, D. M., & Blitstein, J. L. (2003). Methods to reduce the impact of intraclass correlation in group-randomized trials. *Evaluation Review, 27*(1), 79–103. https://doi.org/10.1177/0193841X02239019.

Nation, M., Crusto, C., Wandersman, A., Kumpfor, K. L., Seybolt, D., Morrissy-Kane, E., & Davino, K. (2003). What works in prevention: principles of effective prevention programs. *American Psychologist, 58*(6/7), 449–456. https://doi.org/10.1037/0003-066X.58.6-7.449.

Niolon, P. H., Vivolo-Kantor, A. M., Latzman, N. E., Valle, L. A., Kuoh, H., Burton, T., Taylor, B. G., & Tharp, A. T. (2015). Prevalence of teen dating violence and co-occurring risk factors among middle school youth in high-risk urban communities. *Journal of Adolescent Health, 56*, S5–S13. https://doi.org/10.1016/j.jadohealth.2014.07.019.

*Peskin, M. F., Markham, C. M., Shegog, R., Baumier, E. R., Addy, R. C., & Tortolero, S. R. (2014). Effects of the It's Your Game. Keep It Real Program on dating violence in ethnic-minority middle school youths: a group randomized trial. American Journal of Public Health, 104(8), 1471–1477. DOI: https://doi.org/10.2105/AJPH.2014.301902.

*Peskin, M. F., Markham, C. M., Shegog, R., Baumier, E. R., Addy, R. C., Temple, J. R., Hernandez, B., Cuccaro, P. M., Thiel, M. A., Gabay, E. K., & Tortolero Emery, S. R. (2019). Adolescent dating violence prevention program for early adolescents: the Me & You randomized controlled trial, 2014-2015. American Journal of Public Health, 109(10), 1419–1428. DOI: https://doi.org/10.2105/AJPH.2019.305218.

*Plourde, C., Shore, N., Herrick, P., Morrill, A., Cattabriga, G., Bottino, L., Orme, E., & Stromgren, C. (2015). You the man: theater as bystander education in dating violence. *Arts & Health, 8*(3), 229–247.

*Reidy, D. E., Holland, K. M., Cortina, K., Ball, B., & Rosenbluth, B. (2017). Evaluation of the expect respect support group program: a violence prevention strategy for youth exposed to violence. *Preventive Medicine, 100*, 235–242. https://doi.org/10.1016/j.ypmed.2017.05.003.

*Rizzo, C. J., & Houck, C. (2019). *Partner violence prevention for middle school boys: a dyadic web-based intervention study (project STRONG).* Report submitted to the National Institute of Justice. Retrieved from: https://www.ncjrs.gov/pdffiles1/nij/grants/252517.pdf.

*Rizzo, C. J., Joppa, M., Barker, D., Collibee, C., Zlotnick, C., & Brown, L. K. (2018). Project date SMART: a dating violence (DV) and sexual risk prevention program for adolescent girls with prior DV exposure. Prevention Science, 19(4), 416–426. DOI: https://doi.org/10.1007/s11121-018-0871-z.

Sampert, S. (2010). Let me tell you a story: English-Canadian newspapers and sexual assault myths. *Canadian Journal for Women and the Law, 22*(2), 301–328.

*Sanchez-Cesareo, M. (2002). *Outcome evaluation of the youth project: a school based teen dating violence prevention program*. PhD dissertation. DePaul University, Chicago, Illinois.

*Sanchez-Jimenez, V., Munoz-Fernandez, N., & Ortega-Rivera, J. (2018). Efficacy evaluation of "Dat-e Adolescence": a dating violence prevention program in Spain. PLoS One, 13(10), 1–23. DOI: https://doi.org/10.1371/journal.pone.0205802.

Sanchez-Meca, J., Marin-Martinez, F., & Chacon-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*(4), 448–467. https://doi.org/10.1037/1082-989X.8.4.448.

*Sargent, K. S., Jouriles, E. N., Rosenfield, D., & McDonald, R. (2017). A high school-based evaluation of TakeCARE, a video bystander program to prevent adolescent relationship violence. Journal of Youth and Adolescence, 46(3), 633–643. DOI: https://doi.org/10.1007/s10964-016-0622-z.

*Schramm, D. G., & Gomez-Scott, J. (2012). Merging relationship education and child abuse prevention knowledge: an evaluation of effectiveness with adolescents. *Marriage & Family Review, 48*(8), 792–808.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S D. (1998). *Preventing crime: what works, what doesn't, what's promising*. National Institute of Justice Research in Brief. U.S. Department of Justice. Retrieved from https://www.ncjrs.gov/pdffiles/171676.PDF.

*Silverman, J. G. (1999). *An outcome evaluation of a school-based curriculum for the primary prevention of dating violence among adolescents*. PhD dissertation. The University of Georgia, Athens, Georgia.

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

Simon, T. R., Miller, S., Gorman-Smith, D., Orpinas, P., & Sullivan, T. (2010). Physical dating violence norms and behavior among sixth-grade students from four U.S. sites. *Journal of Early Adolescence, 30*(3), 395–409. https://doi.org/10.1177/0272431609333301.

Smith, D. M., & Donnelly, J. (2008). Adolescent dating violence. *Journal of Prevention & Intervention in the Community, 21*(1), 53–64. https://doi.org/10.1300/J005v21n01_04.

*Southgate, K. (2016). *A mixed methods exploration of the impact of a healthy relationship programme on adolescents in the UK*. PhD dissertation. University of Surrey, Guilford, Surrey, United Kingdom.

Spence, S. H., Sheffield, J. K., & Donovan, C. L. (2003). Preventing adolescent depression: an evaluation of the problem solving for life program. *Journal of Consulting and Clinical Psychology, 71*(1), 3–13. https://doi.org/10.1037/0022-006X.71.1.3.

Spikes, C., & Sternadori, M. (2018). Boomerang effects of sexual-violence prevention messages on college men's attitudes. *Building Healthy Communities Journal, 2*(2), 45–57.

Steichen, T. J. (1998). Tests for publication bias in meta-analysis. *Stata Technical Bulletin, 41*(19), 9–15.

Sterne, J. A. C., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal, 4*(2), 127–141.

Taylor, B. G., & Mumford, E. A. (2016). A national descriptive portrait of adolescent relationship abuse: results from the national survey on teen relationships and intimate violence. *Journal of Interpersonal Violence, 31*(6), 963–988. https://doi.org/10.1177/0886260514564070.

Taylor, K. A., Sullivan, T. N., & Farrell, A. D. (2015). Longitudinal relationships between individual and class norms supporting dating violence and perpetrations of dating violence. *Journal of Youth and Adolescence, 44*, 745–760. https://doi.org/10.1007/s10964-014-0195-7.

Temple, J. R., Le, V. D., Muir, A., Goforth, L., & McElhany, A. L. (2013). The need for school-based teen dating violence prevention. *Journal of Applied Research on Children, 4*(1), 1–11.

Teten Tharp, A., McNaughton Reyes, H. L., Foshee, V., Swahn, M. H., Hall, J. E., & Logan, J. (2017). Examining the prevalence and predictors of injury from adolescent dating violence. *Journal of Aggression, Maltreatment & Trauma, 26*(5), 445–461. https://doi.org/10.1080/10926771.2017.1287145.

Thornberg, R., Wanstrom, L., Hong, J. S., & Espelage, D. L. (2017). Classroom relationship qualities and social-cognitive correlates of defending and passive bystanding in school bullying in Sweden: a multilevel analysis. *Journal of School Psychology, 63*, 49–62. https://doi.org/10.1016/j.jsp.2017.03.002.

Ting, S. R. (2009). Meta-analysis on dating violence prevention among middle and high schools. *Journal of School Violence, 8*(4), 328–337. https://doi.org/10.1080/15388220903130197.

van der Put, C. E., Assink, M., Gubbels, J., & Boekhout van Solinge, N. F. (2018). Identifying effective components of child maltreatment interventions: a meta-analysis. *Clinical Child and Family Psychology Review, 21*, 171–202. https://doi.org/10.1007/s10567-017-0250-5.

Visschers, J., Jaspaert, E., & Vervaeke, G. (2017). Social desirability in intimate partner violence and relationship satisfaction reports: an exploratory analysis. *Journal of Interpersonal Violence, 32*(9), 1401–1420.

What Works Clearinghouse. (2017). *Procedures handbook. Version 4.0.* Retrieved from: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf.

*Wolfe, D. A., Crooks, C., Jaffe, P., Chiodo, D., Hughes, R., Ellis, W., Stitt, L., & Donner, A. (2009). A school-based program to prevent adolescent dating violence. Archives of Pediatrics and Adolescent Medicine, 163(8), 692–699.

Chelsey Lee is a doctoral student in the SFU School of Criminology. Chelsey has spent the last 2 years as the Project Coordinator for a large-scale evaluation of the Metro Vancouver YWCA Youth Education Programs for at-risk students in Surrey and Vancouver. She has also contributed to several meta-analyses focused on interventions to reduce criminal recidivism, including halfway houses and day reporting centers. Her recent work can be found in several scholarly journals such as Crime and Delinquency, Violence Against Women, Journal of Interpersonal Violence, Studies in Higher Education, Journal of Offender Rehabilitation, and the International Journal of Offender Therapy and Comparative Criminology.

Jennifer S. Wong is an Associate Professor in the School of Criminology at Simon Fraser University. Her work employs applied, quantitative, and qualitative methods to study issues of delinquency/crime prevention and intervention, focusing primarily on evaluation and policy analysis in the areas of crime prevention and crime control policy. Dr. Wong's recent work has examined evaluations of intimate partner violence intervention programs, the impact of prior deportation of illegal immigrants on recidivism, the relationship between media framing of homicide and the use of images in newspaper articles, and meta-analyses on gang prevention programs, restorative justice programs for at-risk youth, and intensive supervision programs.