Examining the Effects of Dating Violence Prevention Programs:
A Systematic Review and Meta-Analysis

by

Chelsey Lee

BA (Hons), Simon Fraser University, 2016

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
School of Criminology
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
FALL 2019

Approval

Name:	**Chelsey Lee**
Degree:	**Master of Arts (Criminology)**
Title:	**Examining the effects of dating violence prevention programs: A systematic review and meta-analysis**
Examining Committee:	**Chair:** Bill Glackman Associate Professor
	Jennifer Wong Senior Supervisor Associate Professor
	Sheri Fabian Supervisor University Lecturer
	Kevin Douglas External Examiner Professor Department of Psychology Simon Fraser University
Date Defended/Approved:	December 16, 2019

Abstract

Dating violence is a prevalent issue among adolescents and refers to any physical, psychological, or sexual violence perpetrated by a partner in a close relationship (Centers for Disease Control and Prevention, 2019a). Prevention programs aim to increase awareness of dating violence and promote healthy relationships. This meta-analysis examines the efficacy of programs targeting adolescents at increasing knowledge about dating violence, changing attitudes towards dating violence behaviours, increasing bystander behaviours, and reducing incidents of adolescent dating violence perpetration and victimization. A systematic search yielded 37 studies contributing 71 independent effect sizes. Studies were pooled by outcome measure and results suggest that prevention programs have a significant, positive impact on measures of knowledge, attitudes, and violence perpetration, but did not significantly impact experiences of victimization or bystander behaviours. In addition, nine moderators were used to examine the impacts of program, participant, and study characteristics. Implications and recommendations for future research are discussed.

Keywords: dating violence; adolescents; prevention programs; meta-analysis

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my supervisory committee. To Dr. Jennifer Wong, thank you for going above and beyond for me throughout this entire process; I would not have made it here without you. Thank you for saying “yes” to me as an undergraduate student asking for research opportunities, despite having never met me. That response and subsequent opportunities forever changed the course of my education. I have grown in immeasurable ways due to your guidance, wisdom, and endless support and encouragement. I could not have asked for a better supervisor and mentor and will be forever grateful to you.

To Dr. Sheri Fabian, thank you for being a constant presence throughout the course of my education, from my first ever criminology course to the end of my Master’s degree. As you said, this journey has come full circle. Your guidance and insight will always be appreciated. To Dr. Kevin Douglas, thank you for your invaluable feedback; my research has only benefited from your insightful comments.

To my amazing friends, thank you for your constant understanding when I took days to respond to messages or turned down more hangouts than I accepted, and still allowing me to vent about the stresses of grad school. Your friendship means the world to me and I can’t imagine doing this without your support.

Vanja, Kelsey, and Jessica, no amount of thanks will ever be enough; you three are the best cheerleaders a girl could ever ask for. Thank you for keeping me sane and making this experience bearable. You all had more confidence in me than I did, and I would not have survived the whirlwind that was the writing process without you.

Finally, a huge thank you to my family, without whom this would not have been possible. Thank you for continuously supporting my never-ending education and for putting up with my exhausted, stressed out self (and apologies in advance for another four years of this). I’ve made it this far because of you.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction

Dating violence is a serious and prevalent issue among adolescents, with approximately 43% of reported dating violence incidents in Canada occurring among youth aged 15-24 (Hotton Mahoney, 2010). In the United States, an estimated 1 in 11 female and 1 in 15 male teens have experienced physical violence perpetrated by a dating partner (Centers for Disease Control and Prevention [CDC], 2019a). Dating violence can include a variety of behaviours spanning psychological, physical, and sexual domains and has been found to be associated with substantial consequences (Cascardi, King, Rector, & DelPozzo, 2018; CDC, 2019a; Hickman, Jaycox, & Aranoff, 2004). Consequences such as depression, anxiety, disordered eating, and physical injury all have documented relationships with experiences of dating violence and have evidence of long-term impacts (Ackard & Neumark-Sztainer, 2002; Callahan, Tollman, & Saunders, 2003; Coker et al., 2000; Tharp et al., 2017). Given the serious, long-term consequences that can arise from experiences of dating violence, the prevalence of such behaviours is particularly concerning. Additionally, research has shown that many adults who experience intimate partner violence first experienced such violence in a relationship during adolescence (Hotton Mahoney, 2010). Early intervention and prevention of dating violence behaviours is imperative.

Many adolescent dating violence prevention programs have been developed in recent years, with a heavy focus on school-based programs. Many are delivered in school classrooms (often in health class) during middle or high school (e.g., Safe Relationships Program (Lowe, Jones, & Banks, 2015); Fourth R (Cissner & Hassoun Ayoub, 2014)). Programs aimed at middle and high school students often work to address dating violence by targeting negative attitudes and behaviours, while developing relationship and conflict resolution skills (e.g., Adler-Baeder et al., 2007; Lowe et al., 2015; MacGowan, 1997). Many evaluations of dating violence prevention programs have been done previously, however, the findings have been mixed. Several meta-analytic reviews have attempted to examine the overall effectiveness of dating violence

prevention programs; however, these reviews have disparate results and accompanying limitations.

The current study seeks to expand and improve upon the existing meta-analytic literature regarding adolescent dating violence prevention programs. The purpose of this study is to use systematic review and meta-analysis to synthesize and evaluate the overall effectiveness of dating violence prevention programs at improving adolescent attitudes and behaviours. Specifically, this study examines the impact of dating violence prevention programs at increasing adolescents' knowledge about dating violence, improving attitudes and beliefs towards dating violence behaviours, reducing incidents of dating violence perpetration, reducing experiences of dating violence victimization, and increasing bystander intentions and/or behaviours. This study is distinct from previous research in several ways, such as including both published and unpublished works and a broader range of research designs, providing a more updated set of included studies allowing for the examination of new research, and implementing a more specific definition of dating violence, with stricter selection criteria. Together, these distinctions help this study to provide important insight into the effectiveness of dating violence prevention programs and expand on existing findings.

Chapter 2.

Background on dating violence and meta-analysis

2.1. Review of the literature: Adolescent dating violence and prevention programs

2.1.1. Defining dating violence

Dating violence is a widespread issue and can refer to a variety of behaviours within a non-married dating relationship. Generally, the literature has defined dating violence as the perpetration of physical, emotional, or sexual abuse, stalking, or verbal threats between two people in a dating relationship (Cascardi, King, Rector, & DelPozzo, 2018; CDC, 2019a; Hickman et al., 2004). Dating violence is not defined solely by any one subtype of violence; many behaviours are interrelated and often co-occur. For example, research has found that emotional or psychological violence often precedes physical abuse (Hyden, 1995; Ryan, 1995).

Physical dating violence

Physical dating violence is often the type of violence that first comes to mind when discussing relationship violence. Physical dating violence can include a variety of behaviours, including both direct physical aggression and the threat of physical force (Cornelius & Resseguie, 2006). Examples of physical abuse include pushing, hitting, slapping, hair-pulling, punching, biting, kicking, choking, and beating with an object (Smith & Donnelly, 2000). Often, physical abuse escalates into more serious actions, though they may start off minor; similarly, instances of violence may begin sporadically and escalate into more persistent patterns (Smith & Donnelly, 2000). With respect to the motivations behind physical abuse, it is commonly believed that control is a considerable component with the idea that force is a way to manipulate the partner (Smith & Donnelly, 2000). However, this is not always the case in adolescent dating relationships; instead, these behaviours may be used to gain attention or express interest (Shapiro, Baumeister, & Kessler, 1991). Physical dating violence is one of the most commonly researched forms of violence, however, focusing solely on this category narrows the

definition and can provide biased estimates of prevalence (Cornelius & Resseguie, 2006).

Psychological or emotional dating violence

Psychological or emotional dating violence is the most prevalent category of abuse experienced by adolescents (Hickman et al., 2004; Leen et al. 2012; Niolon et al. 2015; Taylor & Mumford, 2016). Psychological abuse consists of any acts intended to threaten, denigrate, deceit, or control, and which destabilize the dating partner or compromise his/her well-being (Orpinas, Nahapetyan, Song, McNicholas, & Reeves, 2012). Specific behaviours can include a range of emotional and verbal tactics and may include verbal abuse, such as name calling or ridicule, isolation, jealousy or possessiveness, and threats of abuse to a victim or their friends, pets, or possessions (Crowell & Burgess, 1996). Psychological abuse can be particularly harmful for adolescents due to the fact that they are still developing their identities and sense of self; they are more likely to lack self-confidence, and, thus, are less likely to be able to stand up to the emotional abuse (Smith & Donnelly, 2000).

Sexual dating violence

Sexual violence refers to any sexual activity that occurs without consent that is freely given (CDC, 2019b). For example, this can include acts of intimidation or coercion to engage in sexual acts or any activity leading up to intercourse or other sexual acts (Smith & Donnelly, 2001). Sexual violence is the least reported form of dating violence, with rates typically lower than psychological or physical dating violence (Hickman, et al., 2004; Taylor & Mumford, 2016). However, while it may be the least reported form of dating violence, incidents of sexual violence may in fact be occurring at similar rates as physical or psychological violence but suffer a higher rate of underreporting. Given that sexual violence in other contexts is well known to be highly underreported, it would not be unlikely for it to be underreported in this context as well (Rottenburg & Cotter, 2018).

Electronic (cyber) dating violence

An additional type of dating violence is that of electronic or cyber dating violence. As technology becomes more prevalent, it offers new ways for harassment and violence to occur. The CDC (n.d.) defines electronic aggression as any harassment or bullying that occurs via e-mail, a chat room, instant messaging, a website (including blogs), or

text messaging. It follows then, that electronic dating violence refers to any act of electronic aggression in the context of a dating relationship. These behaviours can include posting insulting or threatening messages about a partner online, using a partner's cell phone or text messages to monitor their behaviour, checking a partner's messages without permission, or leaving threatening messages (Drauker & Martsolf, 2010). Technology plays an increasing role in adolescent romantic relationships, with many starting and/or ending their relationship via an electronic device (Baker & Carreno, 2016). However, despite electronic dating violence becoming more prevalent, it is not widely reported on or studied; the majority of research related to dating violence focuses on psychological, physical, or sexual violence (Cutbush, Williams, Miller, Gibbs, & Clinton-Sherrod, 2018). As this type of abuse is rarely measured in existing evaluations of dating violence prevention programs, challenges exist in assessing the impacts of programs on electronic abuse.

Dating violence in the adolescent context

There are several unique elements regarding the context of adolescent dating violence. The first is the adolescent state of development, as adolescents are at a critical point in their cognitive and personal development (Kerig, 2010; Smith & Donnelly, 2000). Numerous psychological changes occur as adolescents begin to build their sense of self and self-confidence (Lapierre, Paradis, Todorov, Blais, & Hebert, 2019). Due to this period of change, the pressure of social norms can carry substantial weight among the adolescent population (Smith & Donnelly, 2000). The new dating experiences that begin at this age can be full of new challenges, which can lead to conflict and stress. This in turn, can lead to an increased likelihood of engaging in behaviours to maintain the relationship, such as perpetrating or tolerating violence (Lapierre et al., 2019; Smith & Donnelly, 2000).

Adolescent dating relationships often differ from their adult counterparts. When referring to adult intimate partner violence, the relationship between those involved tends to be a serious, committed relationship, often common-law or spousal relationships (Fairbairn & Dawson, 2013). In the context of adolescent relationships, however, there is a greater variability in how relationships are defined (Pittman, Wolfe, & Wekerle, 2000). Adolescence is a time for emerging romantic relationships; these relationships often develop in group settings and then become more independent as the adolescents gain

more experience (Ellis & Wolfe, 2015). For example, an adolescent dating relationship may be defined as a single date or a long-term relationship, and there is no consensus regarding the amount of time required to constitute a 'relationship' (Pittman et al., 2000).

Signs of abuse often go unnoticed in adolescent relationships. Due to their lack of experience with romantic relationships, adolescents are less likely to recognize what constitutes abuse (Kerig, 2010; Smith & Donnelly, 2000). Myths regarding issues such as intimate partner violence are widespread; inaccurate perceptions and beliefs are ingrained in society and are widely perpetuated (Sampert, 2010). The prevalence of inaccurate information regarding gender-based violence can negatively impact adolescents' ability to identify specific behaviours as dating violence; they may also mistake certain abusive behaviours for signs of caring (Smith & Donnelly, 2000). If adolescents are primarily exposed to misconceptions about love and healthy relationships, they are more likely to perpetuate these same misconceptions in their own relationships (Lichty & Gowan, 2018). Additionally, adolescents often lack the knowledge and skills to properly address any conflict situations that may arise, thus increasing the risk for abusive behaviours (Kerig, 2010; Smith & Donnelly, 2000). These issues highlight the importance of providing knowledge regarding unhealthy relationships and dating violence through early prevention programs.

2.1.2. Prevalence of dating violence among adolescents

Dating violence among adolescents is highly prevalent in North America across regional and national populations, spanning all forms of violence perpetration and victimization (Taylor & Mumford, 2016). In 2013, Taylor and Mumford (2016) conducted a nationally representative survey in the United States among youth aged 10-18 (only ages 12-18 were reported on) regarding their experiences with dating violence. The total sample included 1,804 youth, 37% (n=667) of whom reported having been in a dating relationship. Of those who reported a recent relationship, 69% reported being victimized by their partner and 63% reported perpetrating violence against their partner. When examined by gender, 69% of boys reported being victimized and 62% reported perpetrating dating violence. Girls reported similar rates, with 69% reporting victimization and 63% reporting perpetration. These high rates are consistent with rates found by other studies at the regional and local levels. For example, Niolon et al. (2015) surveyed middle school students in four cities in the United States; of the participants who

reported having been in a dating relationship, 77% reported perpetrating at least one form of dating violence against their partner. However, specific estimates can vary, and often differ widely based on how dating violence is defined (i.e., emotional, physical, or sexual violence; Hoefer, Black & Ricard, 2015).

When studying the prevalence of dating violence, psychological or emotional violence is not often reported on independently from a global measure of violence. However, when it is reported, psychological violence typically has higher rates compared to physical or sexual violence (Leen et al., 2012). With regard to perpetration, reported estimates range from 62% to 77% (Niolon et al., 2015; Taylor & Mumford, 2016). Psychological abuse victimization rates range from 42% to 66% (Hickman et al., 2004; Taylor & Mumford, 2016).

Physical and sexual violence on the other hand, often have much lower estimates of both perpetration and victimization. Wincentak, Connolly, and Card (2017) conducted a meta-analysis examining the prevalence of teenage dating violence perpetration and victimization. Their analysis included 101 studies and found an overall prevalence rate of approximately 20% for physical dating violence (though the range had a high of 61%). This estimate is consistent with reports that suggest physical violence perpetration rates range from 12% to 33% and victimization rates range from 8% to 25% (Hickman et al., 2004; Kann et al., 2017; Niolon et al., 2015; Taylor & Mumford, 2015).

With regard to sexual dating violence, this type of abuse tends to be the least prevalent of the three types. Wincentak et al. (2017) found an overall rate of approximately 9% based on a series of studies that report on sexual violence perpetration and victimization and provide an overall range of <1% to 54%. Other reports of sexual dating violence perpetration rates have provided estimates of 12% to15% for perpetration and 5% to18% for victimization (Hickman et al., 2004; Kann et al., 2017; Niolon et al., 2015; Taylor & Mumford, 2015).

Gender differences in prevalence rates

The rates of dating violence between males and females tend to vary based on the type of violence measured. Reports of physical dating violence perpetration have been found to differ significantly between males and females, with females being more likely to report being physically violent (an overall rate of 25% for females and 13% for

males; Wincentak et al., 2017). Victimization rates on the other hand are consistent between males and females, with both genders reporting an overall rate of 21% for physical dating violence (Wincentak et al. 2017). Sexual violence shows the opposite pattern of reporting, with males being significantly more likely to report perpetrating sexual violence against their partner than females (10% for boys, 3% for girls; Wincentak et al., 2017). Conversely, sexual violence victimization is reported significantly less often by males than by females (8% of boys, 14% of girls; Wincentak et al., 2017).

The meta-analysis by Wincentak et al. (2017) did not include measures of psychological violence; other research presents mixed results regarding gender differences for this type of abuse. Niolon et al. (2015) found that females reported significantly more emotional/verbal abuse perpetration than did males (82% of females and 72% of males). Additionally, results by Foshee (1996) demonstrate significantly higher rates of psychological abuse victimization among adolescent females. However, Taylor and Mumford (2016) found no significant differences in perpetration or victimization of psychological violence, with approximately 68% of both male and female populations reporting abuse.

Age trends

As many adult victims of intimate partner violence report first being victimized as an adolescent, it is important to intervene early (Hotton Mahoney, 2010). To do so, it is necessary to understand the patterns of perpetration and victimization with respect to age. Bonomi et al. (2012) investigated age of onset of dating violence and patterns of abusive experiences using a retroactive design with college-aged students. Participants were asked to report on psychological, physical, and sexual dating violence experiences, concentrating on age of first experience, number of abusive partners, and number of occurrences for each type. Results found that 44.7% of females first experienced psychological violence (e.g., controlling behaviour) between the ages of 13 and 15, and 62.5% first experienced pressured sex between age 16 and 17. Males show a more consistent age of onset across the three types of violence, with 16-30% of victimization occurring before the age of 15. These results are similar to those found more recently by Shorey et al. (2017). Shorey and colleagues conducted a six-year longitudinal study of high school students regarding age patterns of dating violence and found that the risk of dating violence was greatest for females between the ages of 15-16, with risk increasing

in later adolescence. Males had a relatively consistent level of risk throughout adolescence, with a slight increase in risk between ages 18-20. The results of these two studies suggest that victimization patterns for females vary by age and type of violence. Additionally, they are most likely to experience their first incident between the ages of 13 and 16, which is consistent with when the risk of victimization is greatest. Conversely, males appear to have more consistent patterns of onset across abuse types and persistence throughout adolescence.

Prevalence across geographic location

The prevalence rates discussed so far have primarily been based on studies from the United States; however, these rates appear to be consistent with Canadian and Western European countries as well. In 2008, 43% of police-reported incidents of dating violence in Canada occurred among those aged 15-24 years (Hotton Mahoney, 2010), which is similar to the rates found previously by Wolfe et al., (2001). Wolfe and colleagues surveyed Ontario youth and found that 28% of boys and 19% of girls reported physical dating violence victimization, and 36% of boys and 43% of girls reported sexual dating violence victimization. The rates found in North America are also consistent with rates found in Western Europe and the UK (Leen et al., 2012). Studies from North America, Europe, and the UK have all shown consistent rates for physical violence, with a range of 10%-20%, while sexual violence victimization has a range of 17% to 23% for girls and 3.8% to 6.6% for boys (Brzank, Liepe, Schillmoller, & Blattner, 2014; Leen et al., 2012). While there is some variation between geographic regions, the overall estimates appear quite similar.

2.1.3. Consequences of dating violence

The high prevalence of dating violence is especially concerning given the potential consequences that can arise. Research has found significant correlational links between experiences of dating violence and negative psychological, behavioural, and physical outcomes (e.g., Ackard & Neumark-Sztainer, 2002; Banyard & Cross, 2008; Callahan et al., 2003; Coker et al., 2000; Tharp et al., 2017). The specific consequences are discussed in detail in the following sections.

Impacts of dating violence on psychological well-being

Given that adolescents are at a critical point in the development of their personal identity and sense of self, dating violence victimization has a considerable potential to impact their overall psychological well-being and development (Callahan et al., 2003). As such, understanding the possible impacts of dating violence is useful for developing appropriate resources for addressing dating violence. Callahan et al. (2003) surveyed 190 high school students (224 prior to exclusions) regarding their experiences with physical and sexual dating violence, and possible outcomes relating to physical injury and elements of psychological well-being such as anxiety, depression, self-esteem, life satisfaction, and post-traumatic stress symptoms. Overall, results suggest that experiences of dating violence are significantly related to reports of lower psychological well-being for both boys and girls. Results show a significant relationship between dating violence victimization and post-traumatic stress for girls. With respect to boys, while they reported fewer instances of dating violence victimization, those who did report victimization experiences showed very similar outcomes to victimized girls. These outcomes include reports of higher levels of depression, anxiety, and post-traumatic stress (Callahan et al., 2003). Additionally, Banyard et al. (2008) and Foshee, McNaughton Reyes, Gottfredson, Chang, and Ennett (2013) also found dating violence victimization to be related to negative mental health issues, such as experiences of depression for boys and girls. Psychological impacts are also known to be long lasting, with consequences persisting at six-month follow-up measurements (Brown et al., 2009). These results suggest that adolescents' psychological well-being can be significantly impacted by physical and sexual dating violence victimization and effective prevention efforts are necessary.

When examining specific types of dating violence, there are also significant differences in psychological impacts based on the type of violence experienced (Foshee et al., 2013). While physical abuse was more significantly related to behavioural outcomes, psychological abuse was significantly related to internalizing symptoms, including feelings of depression and anxiety (Foshee et al., 2013). Additionally, experiences of sexual dating violence is significantly related to increased instances of suicidal ideation and suicide attempts, which is particularly concerning (Ackard & Neumark-Sztainer, 2002; Coker et al., 2000). After controlling for factors such as age, race, and other abuse by an adult, adolescents who had experienced dating violence

and rape were significantly more likely to experience suicidal thoughts or attempts than their non-abused peers (Ackard & Neumark-Sztainer, 2002). These results highlight the importance of addressing all forms of dating violence in prevention programs.

Behavioural consequences of dating violence

Dating violence victimization is also associated with multiple negative behavioural effects. There is evidence that physical and sexual dating violence are associated with increased disordered eating, substance use, and risky sexual behaviours (Ackard & Neumark-Stztainer, 2002; Coker et al., 2000; Silverman, Raj, Mucci, & Hathaway, 2001). Ackard and Neumark-Sztainer (2002) surveyed over 80,000 students in grades 9 and 12 regarding their experiences of dating violence and found that those who experienced dating violence and rape were significantly more likely to use laxatives or diet pills, vomit, and binge eat than their non-abused peers. This finding was consistent with those of Silverman et al. (2001), who also found that girls who had experienced physical or sexual dating violence were more likely to engage in unhealthy weight loss behaviours, such as the use of laxatives or vomiting.

Dating violence is also significantly related to substance use; those who have experienced dating violence are more likely to be substance dependent (Brown et al., 2009). Risky sexual behaviours (e.g., first intercourse before age 15, having multiple sexual partners, engaging in substance use prior to intercourse) and adolescent pregnancy are also significantly correlated with physical and sexual dating violence victimization among adolescent girls; those who have been victimized are significantly more likely to engage in these risky behaviours or report having been pregnant compared to their non-abused peers (Silverman et al. 2001; Silverman, Raj, & Clements, 2004).

Physical consequences of dating violence

Research regarding the specifics surrounding physical injuries sustained through adolescent dating violence is scarce; however, there is evidence that dating violence can result in physical injury ranging from minor to severe. Tharp et al. (2017) examined the prevalence of dating violence-related injuries and medical help-seeking behaviours in youth grades 8 through 12. Results show that 54% of those who reported experiencing dating violence had sustained an injury as a result of the abuse. Physical injuries were

significantly more prevalent in females than males, with scratches and bruises representing the most common injuries reported, though 8.4% of the abused sample reported sustaining injuries serious enough to warrant medical care. These results are consistent with previous findings, in which females reported significantly more injuries than males (70% of females and 52% of males reported injury) and approximately 8 to 9% of adolescents overall reported needing to visit an emergency room due to an injury received via dating violence (Foshee et al., 1996).

2.1.4. Risk-factors for adolescent dating violence

Previous violence

When addressing the topic of dating violence prevention, it is also important to identify the relevant risk factors and possible predictors of dating violence behaviours. One such factor is bullying behaviour, which is a significant predictor of both dating violence perpetration and victimization (Ellis & Wolfe, 2015). Ellis and Wolfe (2015) surveyed Canadian students in grades 9, 10, and 11 regarding their bullying behaviours as well as measures of dating violence. The results found a significant, predictive relationship between reported bullying perpetration and both dating violence perpetration and victimization. These findings are consistent with those from Foshee et al. (2014), whose longitudinal study demonstrated a significant relationship between direct bullying behaviours (e.g., hitting, slapping, or picking on another student) in the sixth grade and perpetration of physical dating violence in the eighth grade.

Similarly, there is a significant predictive relationship between experiencing violence at a young age (e.g., witnessing or experiencing parental violence) and dating violence in adolescence (Foshee et al., 2004; Giordano, Kaufman, Manning, & Longmore, 2015; Latzman, Vivolo-Kantor, Niolon, Ghazarian, 2015). Specifically, those who are exposed to intimate partner violence are more likely to report perpetrating relational aggression later in life (Latzman et al., 2015). Additionally, physical abuse by a parent is a significant predictor of dating violence perpetration (Giordano et al., 2015). With respect to victimization, a family history of violence is also significantly related to an increased risk for adolescent girls being psychologically and/or physically/sexually victimized in their dating relationships (Vezina et al., 2015). These results, in combination with the findings regarding the predictive power of bullying, are consistent

with the cycle of violence concept: those who experience violence are more likely to experience and/or perpetrate violence (Widom, 1989).

Peer associations

Friends and peer associations are also significantly related to dating violence behaviours. Violence is arguably a learned behaviour and can be heavily influenced by peers. This is particularly true in adolescence as peers begin to take on a significant role in an adolescent's life (Connolly, Furman, & Konarski, 2000). Giordano et al. (2015) examined the impact of friendships and school context on dating violence behaviours and found friend violence to be a significant predictor of dating violence perpetration. Those who were victimized by their friends were more likely to perpetrate dating violence behaviours (Giordano et al., 2015). Similar results were found by Garthe, Sullivan, and McDaniel, (2017), with peer victimization being significantly related to dating violence. Garthe et al. (2017) suggest that those who are victimized by their peers are more likely to feel isolated, which in turn can lead to more aggressive or negative behaviours. Garthe and colleagues also found a significant relationship between aggressive peers and dating violence perpetration; this may be due to an increased likelihood of associating with similar individuals. In sum, if an adolescent's peers are aggressive, it is more likely the adolescent will be aggressive as well. Additionally, level of violence within a school is also a significant predictor of dating violence behaviours, research finds that those attending a school with a mid-range level of violence have higher rates of dating violence perpetration compared to those attending a low-violence school (Giordano et al., 2015). These results illustrate the impact the peer environment can have on adolescents.

Substance abuse

Additionally, substance use has been identified as a significant predictor of dating violence, with those who engage in substance use more likely to engage in dating violence than those who do not use illegal substances (Chase, Treboux, & O'Leary, 2002; Vagi et al., 2013). For instance, Temple, Shorey, Fite, Stuart, and Le (2013) found that those youth who abused drugs and alcohol were significantly more likely to perpetrate physical dating violence behaviours one year later. Reyes, Foshee, Tharp, Ennett, and Bauer (2015) found similar results, with substance-involved youth more likely to engage in dating violence. However, it is important to note that these results

differ based on the social and situational context. For example, strong neighbourhood control buffers the relationship between physical dating violence perpetration and substance abuse; this may be due to increased availability of prosocial supports or greater exposure to prosocial conflict management, which also impact the likelihood of dating aggression (Reyes et al., 2015). Conversely, family violence exacerbates the relationship between substance abuse and dating violence, with those exposed to family violence being more likely to engage in abusive behaviours (Reyes et al., 2015).

Depression

Depression, anxiety, and low self-esteem have all been found to be significant risk factors for dating violence. Foshee et al. (2004) conducted a longitudinal study with a sample of 8th and 9th grade students in rural North Carolina to investigate a series of possible predictors of physical and sexual dating violence. Results found depression to be a significant risk factor for females with respect to the onset and continued victimization of sexual violence. Lehrer, Buka, Gortmaker, and Shrier (2006) report similar findings, with depressive symptoms being significantly associated with an increased risk for partner violence among adolescents. These results suggest that targeting adolescent mental health may be beneficial in preventing dating violence.

Numerous risk factors for adolescent dating violence perpetration and victimization have been identified throughout the literature (e.g., mental health problems, aggressive beliefs, poor family quality, peer influence, bullying behaviours; Connolly et al., 2000; Ellis & Wolfe, 2015; Foshee et al., 2004; Vagi et al., 2013). However, it is important to note that causality cannot be inferred from any of these factors discussed. While several significant relationships have been identified, the directionality of the relationships is unknown. In an attempt to address this issue, Vagi et al. (2013) specifically examined studies that established temporal order of risk and predictive factors in relation to adolescent dating violence. A total of 19 studies discussing risk factors were identified, resulting in a total of 53 risk factors. The risk factors include mental health problems, youth violence, substance use, aggressive thoughts, and risky sexual behaviours, among others. However, while Vagi et al., (2013) found a considerable number of risk factors with temporal order, they also excluded numerous other studies focused on similar risk factors simply because there was no temporal order. This illustrates that many risk factors are not unidirectional. This is further

evidenced by the fact that several of the consequences of dating violence discussed in section 2.1.3 have also been identified here as risk factors or predictors of violence (e.g., depression, anxiety, substance abuse). Whether the factors are predictors of violence, consequences, or both, they are nonetheless harmful to adolescents and are important to consider in the context of preventing dating violence behaviours.

2.1.5. Adolescent dating violence prevention efforts

Legislation and policy

Current efforts to prevent adolescent dating violence include the use of government legislation or policy. In the United States, many states have legislation relating to domestic violence, however, many are primarily targeted towards adults. While adolescents are afforded certain rights by some laws (e.g., filing for a protection order), many are not applicable to adolescents (Hoefer, Black, & Ricard, 2015). With regard to adolescents specifically, approaches differ widely by state; for example, some states require schools to offer dating violence prevention programming and policy (e.g., Nebraska, Ohio), while other states require schools to implement dating violence education but do not require a policy for addressing cases of dating violence (e.g., New Jersey; Black, Hawley, Hoefer, & Barnett, 2017; Break the Cycle, 2010).

Dating violence policy is not universally applied, with many schools having policies only relevant to bullying or other forms of harassment (Cascardi et al., 2018). While bullying and other harassment behaviours are related to dating violence, with a considerable amount of overlap in behaviours, having policy with overly broad definitions of behaviours can be problematic (Cascardi, Brown, Iannarone, & Cardona, 2014; Cascardi et al., 2018). Broad policies can lead to the inappropriate application of such policies, with fragmented or inconsistent implementation due to the lack of specificity (Cascardi et al., 2018; Cornell & Limber, 2015). On the other hand, policies that enable youth to access help and address their situation can be effective at addressing dating violence prevalence (Hoefer et al., 2015). For example, when policies relating to civil protection orders are developed for adolescents and make obtaining a protection order easily accessible, prevalence rates are positively impacted (Hoefer et al., 2015). This suggests that having dating violence-specific policy and law for adolescents may be helpful at effectively reducing rates of adolescent dating violence. However, policy is

only one part of a comprehensive approach to ending and preventing dating violence. It is important to also implement education programming to ensure adolescents have the knowledge and skills needed to address their specific situations.

Dating violence prevention programming

The second component to a comprehensive approach in addressing adolescent dating violence is prevention programming. This includes primary, secondary, and selective interventions (Cornelius & Ressaguie, 2006; Crooks et al., 2019). Primary interventions refer to those programs that can be universally implemented with all adolescent populations and attempt to reduce the occurrence of new dating violence behaviours (Cornelius & Ressaguie, 2006; Crooks et al., 2019). Secondary programs work to address existing violence and behaviours (Cornelius & Ressaguie, 2006), while selective programming targets specific populations who may be at an increased risk for violence (Crooks et al., 2019).

The specific goals and purposes of individual programs vary; however, most programs ultimately aim to end dating violence and promote equitable and respectful relationships among adolescents (Crooks et al., 2019). This is typically done through methods 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 behaviours and reduce/eliminate aggressive behaviours, and increase and develop skills needed to create positive relationships (Crooks et al., 2019).

Generally, adolescent dating violence prevention programs are often administered in-person, in middle and high schools. A school-based setting provides a number of advantages; it is often a convenient location to administer the program. Program organizers are able to reach a large number of youth, space to conduct the program is typically easily available, there are multiple options for program facilitators (teachers and other school staff, such as counsellors), and because school staff can be trained in the curriculum, school-wide change is more sustainable (Temple, Le, Muir, Goforth, & McElhany, 2013). There are also a few limitations of school-based programs, such as excluding vulnerable and at-risk populations (e.g., school drop-outs) and not targeting those who are at greater risk of violence. However, providing education and prevention programming in schools also allows for early intervention, which is important

for preventing future violence; as such, many programs target adolescents between the ages of 11 and 19 (Black et al., 2017; Crooks et al., 2019). Overall, school-based delivery is a common and beneficial method of administration, with many programs following this model (e.g., Safe Relationships Program (Lowe et al., 2015); Fourth R (Cissner & Ayoub, 2014)).

As evidenced by the systematic search results (discussed in Chapter 4), there have been a substantial number of evaluations of existing dating violence prevention programs; however, findings are varied. Generally, when examining outcomes related to knowledge about and attitudes towards dating violence, results are typically positive, though significance levels vary. For example, DeGannes (2009) evaluated a program called “Project Awareness” and did not find any statistically significant results regarding knowledge or attitudes, though the direction of the effect indicated improvement. Conversely, Gardner, Glese, and Parrott (2004) did find significant improvements in knowledge, with the treatment group scoring significantly higher in their evaluation of the “Connections: Relationships and Marriage” program. There is similar variability among studies with respect to behavioural outcomes (e.g., violence perpetration or victimization), though the direction of the effect can also vary. For example, Wolfe et al. (2009) found a negative impact of the “Fourth R” program on incidents of physical violence victimization, while Levesque, Johnson, Welch, Prochaska, and Paiva (2016) found a positive impact of the “Teen Choices” program on physical violence victimization. Given the disparity among study findings, a meta-analytic review is important for understanding the overall landscape of dating violence prevention program efficacy.

2.1.6. Previous reviews of dating violence prevention programs

A literature search identified several previous meta-analytic reviews of dating violence prevention programs: four regarding school-based, adolescent dating violence prevention (de la Rue, Polanin, Espelage, & Pigott, 2017; Edwards & Hinsz, 2014; Fellmeth, Heffernan, Nurse, Hibibla, & Sethi, 2015; and Ting, 2008), and four regarding sexual violence prevention programs (Anderson & Whiston, 2005; Brecklin & Forde, 2001; Flores & Hartlaub, 1998; and Jouriles, Krauss, Vu, Banyard, & McDonald, 2018). While the four studies focused on sexual violence prevention include some measures that overlap with the current study (e.g., rape myth acceptance), the goals and

descriptions of the included programs and studies were deemed to be dissimilar to those of the current study and will not be described in detail. Instead, the focus of this summary is on the four meta-analyses examining adolescent dating violence prevention programs. An overview of these four studies is presented in Table 2.1 and will be discussed in detail below, illustrating how the present study will contribute to the field of dating violence prevention program efficacy.

Table 2-1. Summary of previous meta-analyses of dating violence prevention programs

Study	Inclusion Criteria	# studies included	# studies that overlap with present analysis	Conclusions
de la Rue (2017)	2-group randomized or quasi-experimental design; limited to middle or high school age and setting; includes sexual violence programs, published prior to or in 2013	23	6	Positive effect on knowledge and attitudes; no effect on perpetration or victimization outcomes
Edwards (2014)	Randomized or quasi-experimental design; limited to middle and high school age; includes sexual violence programs; published prior to or in 2011	7	1	Positive effect on a combined measure of dating violence
Fellmeth (2015)	Randomized or quasi-experimental design; middle, high school, and college programs; includes sexual violence programs; published prior to or in 2012	33	2	No significant effects on outcomes of knowledge, attitudes, behaviours, or skills
Ting (2009)	2-group designs; limited to middle or high school age; includes sexual violence programs; published 1990-2007	13	3	Positive effect on outcomes of knowledge and attitudes

Firstly, Ting (2009) conducted a meta-analysis focused on dating violence prevention programs in middle and high schools. Selection criteria resulted in the inclusion of studies with outcomes measuring knowledge of and attitudes towards dating violence and studies that also addressed sexual violence; however, only peer-reviewed sources and official reports were included, thus excluding all theses, dissertations,

books, unpublished manuscripts, and conference papers. These criteria resulted in 13 studies being included in the analyses; the findings suggest an overall improvement in knowledge and attitudes regarding dating violence.

Edwards and Hinsz (2014) also conducted a meta-analysis of school-based prevention programs designed to reduce teen dating violence; their review included seven eligible studies and included outcomes focused on attitudes towards sexual and dating violence, as well as self- and school-reported physical and sexual violence perpetration. Edwards and Hinsz (2014) defined eligible interventions as those "designed to reduce dating and/or sexual violence" (pg. 2). Analyses were conducted using product-moment correlation coefficients, and the authors found an overall positive effect across all outcomes. Heterogeneity was identified among the set of studies; however, moderator analyses were not performed due to the small sample size.

The third of the existing meta-analyses was conducted by Fellmeth et al. (2015) and examined the efficacy of educational and skills-based interventions in preventing relationship violence. This study differed from the two described above, as the inclusion criteria did not limit the target population to middle or high school aged students. College and adult populations were also included, resulting in an analytic sample of 33 independent studies. Analyses addressed outcomes such as victimization of relationship and dating violence, self-reported improvement in mental well-being, and improvements in behaviour and knowledge about dating violence. Results showed no significant effect for all outcomes except for knowledge of dating violence. However, positive effect was significantly heterogenous; the researchers subsequently excluded sensitive studies, which led to a null effect.

The final meta-analysis identified was focused on school-based prevention programs designed to reduce teen dating violence (de la Rue et al., 2017). Analyses included 23 separate studies and analyzed outcome measures such as knowledge of and attitudes towards teen dating violence and various perpetration and victimization behaviours (physical, verbal, and sexual aggression). In addition to programs that aimed to prevent or reduce dating violence, de la Rue and colleagues also included studies that focused on preventing or reducing sexual violence specifically (e.g., rape or sexual coercion). The authors limited the included studies to those that implemented an experimental or quasi-experimental design with a control group, those that were

implemented in schools only (i.e., the authors excluded programs that occurred a community centers or other non-school-based settings). Findings showed mixed results, with a positive effect for knowledge and attitudes towards dating violence, but null or very small effects on perpetration and victimization reports.

The current study aims to extend and improve upon the existing findings of these comprehensive reviews through several key differences. The first is in the definition of the primary programs of interest. The previous meta-analyses in this field all included programs that also target sexual violence (e.g., rape, sexual coercion, date/acquaintance rape); we contend that while sexual violence is a key characteristic of dating violence, it is only one component. Sexual violence can occur in the context of a dating relationship, however it can also occur in non-dating contexts (e.g., stranger, acquaintance). As such, when referring to sexual violence, it might not necessarily be in the context of dating violence. By only addressing sexual violence, a program would not be targeting dating violence as a whole by excluding non-sexual physical and psychological violence. By specifically focusing on programs that address dating violence overall, the current study can examine the impacts of dating violence prevention programs alone, without conflating them with sexual violence-specific programs.

Additionally, previous meta-analyses of dating violence prevention programs have limited the included studies to those with two-group, experimental or quasi-experimental designs and exclude post-test only and single group, pre-test post-test designs. This is a common criterion in meta-analysis to ensure the quality of included studies is strong and the measured pooled effect is representative; designs without a control or matched comparison group may overestimate the treatment effect, resulting in a biased estimate (Carlson & Schmidt, 1999; Lipsey & Wilson, 2001). However, this criterion also excludes a considerable number of studies in the field of dating violence prevention programs. Due to logistical constraints when dealing with an adolescent, school-based population, a randomized control trial or design with a matched comparison group is not always possible. By excluding those studies with ‘weaker’ designs, the conclusions that can be drawn from the field as a whole are limited. While there are limitations to single group designs, there are methodological techniques that can mitigate the concerns (e.g., transforming the effect size to a raw metric; Morris & DeShon, 2002). We contend that the inclusion of single-group, pre-test post-test designs

allows for a more thorough examination of the field of dating violence prevention programs and a more comprehensive pooled effect.

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, organizational reports). This tactic provides a more comprehensive view of the field, as well as minimizes potential publication bias. Finally, this meta-analysis also provides an updated examination of dating violence prevention programs. Although three of the previous studies have been published within the last five years (2014 or later), the searches for the included studies were done considerably earlier, with the most recent of the three being conducted in 2013 (de la Rue et al, 2017). While this was only six years prior to the time of the present literature search, the present search identified 57% of the included studies as being published in 2014 or later. Overall, by using a more current literature search, a more specific definition of dating violence prevention, including additional research designs, and including unpublished works, the current study has very little overlap with the included studies of previous meta-analyses. Given the increase in relevant research, this meta-analysis provides updated, comprehensive results.

2.2. Review of the literature: Systematic review and meta-analysis

2.2.1. Systematic review and meta-analysis as methods of research synthesis

Meta-analysis is a quantitative method of research synthesis, in which existing research is combined and summarized (Lipsey & Wilson, 2001). There are several different methods of synthesizing existing research, each with their own strengths and weaknesses. Card (2011) describes the methods as situated on a continuum of qualitative and quantitative methods. On one side are the qualitative forms of research synthesis, such as narrative literature reviews. In a narrative literature review, the reviewer qualitatively discusses existing studies and draws conclusions based on the overall findings. However, there is no systematic technique for how the researcher identifies and selects the studies or how they reach their conclusions, resulting in substantial subjectivity (Card, 2011). On the other side of the continuum are the

quantitative forms of research synthesis. In the middle of the continuum is one type of quantitative research synthesis known as ‘vote-counting’. Vote counting is a form of a narrative literature review that draws conclusions based on the significance levels of the primary findings of each study. For example, if a majority of the studies reviewed have positive, statistically significant results, a positive conclusion would be drawn. However, the vote counting approach has several limitations, including that it does not take sample size or magnitude of effect into account when forming conclusions (Bushman & Wang, 2009).

Systematic review

On the quantitative side of the continuum are systematic literature review and meta-analysis. What differentiates a systematic literature review from a narrative review is the pre-determined set of criteria used to conduct a comprehensive search in an attempt to identify all empirical evidence relevant to a specific topic (Card, 2011; Lasserson, Thomas, & Higgins, 2019; Lipsey & Wilson, 2002). Systematic reviews are useful in providing a thorough, up-to-date understanding of existing research; as such it is imperative that the search be comprehensive and exhaustive (Lasserson et al., 2019). To do this, clearly identifying the scope of the search and applicable criteria is necessary not only to ensure comprehensiveness, but also to increase the validity and replicability of findings.

A systematic search involves several components to ensure any and all relevant studies are identified, and inclusion and exclusion decisions are made with a set of *a priori* criteria to ensure transparency and validity (Card, 2011; Lipsey & Wilson, 2001). These pre-determined criteria are used to extract and code detailed data from the included studies. To adequately assess the results of any included studies, a sufficient amount of information is required. The specific information extracted and coded for will vary on the specific topic of research and overall project goals, however Lipsey and Wilson (2002) suggest a set of guidelines and recommend including characteristics of publication, study design, intervention components, participants, and outcomes.

Meta-analysis

Meta-analysis is a quantitative literature review method in which the outcomes of existing research are synthesized and analyzed using statistical methods (Card, 2011;

Lipsey & Wilson 2001). Unlike narrative reviews or reviews which employ vote-counting methods, meta-analysis uses an objective approach that considers both direction and magnitude of effect, rather than only levels of significance. Meta-analysis focuses on the outcomes of a set of studies, rather than the individual study authors' conclusions, to calculate a pooled estimate of the overall treatment effect (Card, 2011). While systematic reviews can be conducted without also conducting a meta-analysis, the combination of the thorough and replicable process of systematic literature review and the statistical techniques for synthesis and analysis work to maximize objectivity and minimize bias (Lipsey & Wilson, 2001).

There are several advantages to using meta-analysis over other methods of research synthesis. The first is that it is transparent and objective. As discussed, systematic review and meta-analysis require the entire process to be decided *a priori*, with a set strategy and explicit criteria for study identification and eligibility (Lipsey & Wilson, 2001). Other methods such as narrative reviews typically do not have a rigorous set of criteria, which results in considerable subjectivity regarding the studies chosen for inclusion, as well as the conclusions drawn (Card, 2011). Determining criteria *a priori* reduces bias and increases objectivity. A pre-determined protocol ensures all relevant criteria are considered throughout the process, thus holding the researcher accountable and increasing the validity of the study.

Meta-analysis is also advantageous for handling information from a large number of studies (Lipsey & Wilson, 2001). While methods such as narrative review are sometimes adequate for small amounts of data, it becomes increasingly difficult to organize and manage the information as the number of studies increases. Similarly, if the set of studies is complex or heterogenous (e.g., if multiple outcomes are presented with differing results), it can be difficult to accurately determine the overall findings (Lipsey & Wilson, 2001). The systematic coding process used in meta-analytic methods can address this issue and manage complex findings and large amounts of data more efficiently. Meta-analysis also allows for subgroup analysis to investigate potential differences in effect magnitude pertaining to participant, intervention, or study characteristics (Lipsey & Wilson, 2001).

In addition, meta-analysis enables greater precision in estimating the treatment effect. Rather than relying only on the statistical significance or the direction of an effect,

meta-analysis can account for both direction and magnitude of effect (Lipsey & Wilson, 2001). As smaller sample sizes require a larger magnitude of effect to attain statistical significance, it is less likely that small sample sizes will result in statistically significant findings; thus, the true effect of these studies may not contribute to the overall conclusion. However, meta-analysis can address this issue by providing greater weight to larger studies (which have smaller amounts of sampling error). This increases the overall power for the analysis, which increases the likelihood of finding meaningful effects should a true effect exist (Lipsey & Wilson, 2001).

2.2.2. Criticisms of meta-analysis

One of the primary criticisms of meta-analysis is the likelihood of dissimilar studies being pooled together in the analysis. Lipsey and Wilson (2001) refer to this as the issue of comparing “apples and oranges” (p. 2). It is imperative that the included studies are measuring the same relationship so they can be meaningfully compared. A pooled effect combining studies that differ across intervention or outcome measure would not be useful. However, particularly in evaluation research in the social sciences, exact replication among studies is rare. It is much more likely for the overall concept to be similar but studies to vary in the specific measures or interventions used (Lipsey & Wilson, 2001). The application of strict selection and coding criteria can work to address this potential limitation; by ensuring that all included studies adhere to the same criteria the commensurability of different interventions and studies can be maximized.

A second criticism of meta-analysis is the issue of publication bias. Publication bias, also known as the ‘file drawer problem’, refers to the trend that only studies with significant findings are published and available (Lipsey & Wilson, 2001; Sterne & Harbord, 2004; Sutton, 2009). Studies with null findings tend to be deemed less important or worthy of publication, with many never being disseminated publicly. This can be problematic for meta-analysis; if only studies with significant effects are identifiable through literature searches, biased conclusions may result from the pooling of the available set of studies.

Additionally, issues regarding the quality of included studies can be a limitation of meta-analysis. Given that meta-analysis involves summarizing existing research, it is limited by the studies that are available. In primary research, if there are any issues with

the raw data the study conclusions may be flawed. The same applies to meta-analysis; the raw data for meta-analysis consists of the individual study findings; if the individual studies have flawed designs and outcomes, the conclusions of the meta-analysis will likewise be flawed. As such, the outcomes of the meta-analysis are only as good as the studies included.

Chapter 3.

Methodology

3.1. Selection criteria

The present study employs systematic review and meta-analysis techniques to examine the effectiveness of dating violence prevention programs. The selection criteria were determined *a priori* and were based on the recommendations of Lipsey and Wilson (2001). While specific eligibility criteria will vary based on the specific topic of the systematic review and meta-analysis, Lipsey and Wilson (2001) recommend considering characteristics in the following categories: 1) distinguishing features of a qualifying study (e.g., the nature of the intervention); 2) research respondents (e.g., age, ethnicity); 3) key variables (e.g., outcomes measured); 4) research designs (e.g., was randomization included); 5) cultural and linguistic range (e.g., location and language of a study); 6) time frame (e.g., when was the study conducted); and 7) publication type (e.g., what types of reports are appropriate for the meta-analysis). The pre-determined inclusion and exclusion criteria are described in the following sections (summarized in Table 3.1).

3.1.1. Inclusion criteria

Types of programs

Eligible studies for inclusion included those focused on dating violence prevention and education. Programs were primarily geared towards increasing education and awareness on the overall topic of dating violence. Specific program objectives included providing participants with information regarding the characteristics, features, and consequences of dating violence and unhealthy relationships, changing attitudes towards dating violence and relationship stereotypes, and/or encouraging healthy relationship behaviours.

Types of participants

Participants were primarily adolescents under the age of 18, who attended middle or high school. There was no specific minimum or maximum age, however programs must have been targeting a predominantly adolescent population.

Types of outcomes

Eligible studies needed to include at least one quantitative outcome measure relevant to knowledge about dating violence, attitudes towards dating violence, dating violence victimization or perpetration, or bystander behaviours. The study must also have provided sufficient data and information to enable the calculation of an effect size.

Types of studies

Included studies used a control or comparison group and/or a pre-test-post-test design. 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.

Sample size

To be included, studies must have involved a sample of at least 20 participants. It is well known that small sample sizes present methodological limitations, such as potential inaccuracies due to greater sampling error (Card, 2011; Lipsey & Wilson, 2001). Though most guidelines for conducting meta-analysis do not specify rules regarding a minimum sample size for inclusion in a meta-analysis, a requirement of at least 20 participants was used to minimize the influence of small sample size limitations.

Language of publication

Included studies were published in English[1].

Location of study

Studies were included if they were conducted in North America or Western Europe[2]. To ensure appropriate comparisons and maximize generalizability, the location was limited to Western countries that are reasonably similar in terms of general culture and perspectives regarding gender-based violence.

[1] Studies published in French were also eligible to be included; however, no French studies were identified during the search process.

[2] Studies conducted in Australia and New Zealand were also eligible for inclusion; however, no studies from these locations were identified during the search process.

3.1.2. Exclusion criteria

Types of programs

Dating violence is part of the broader category of gender-based violence and is a type of intimate partner violence that encompasses acts of physical, psychological, and sexual violence that occurs within the context of a dating relationship (CDC, 2019a). As such, 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 solely focused on sexual violence (e.g., sexual assault, rape), general school violence, non-intimate partner peer violence, and bullying. Although many violence prevention programs exist and may include a component or measure relevant to one of the characteristics or included behaviours of dating violence, the prevention of dating violence may not be the primary aim of the program. The present study aims to assess the effectiveness of programs that specifically target dating violence as the primary goal.

This distinction was made to ensure commensurability between programs. The individual types of violence mentioned (e.g., sexual violence, peer violence, bullying) can occur in contexts other than dating relationships, which will impact the overall program goals and the information presented. When examining prevalence rates, it is evident that many who experience dating violence typically experience more than one type of violence (psychological, physical, or sexual). As such, to fully address dating violence, multiple forms of violence should be included in the program. Dating violence prevention programs are unique in this respect, whereas sexual assault prevention programs for example, typically focus only on sexual assault given the nature of the violence. Thus, it can be argued that dating violence prevention programs and other targeted violence programs are not measuring the same treatment.

Types of participants

Studies were excluded if the target population was college-aged or adults. Additionally, studies were 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, or youths living in a residential facility). The purpose of these exclusions

was to maximize similarities among study populations, and in turn maximize the robustness and generalizability of the results.

Types of outcome measures

As this meta-analysis was focused on knowledge and behaviours relevant to dating violence, studies that only presented outcome measures that were unrelated to this topic (e.g., non-intimate partner peer violence, bullying) were excluded, as were those presenting solely qualitative measures (due to the inability to calculate effect sizes).

Types of studies

Studies were excluded if the research design was a post-test only design, due to the inability to measure change related to the intervention. Additionally, studies were excluded if the comparison group was deemed inappropriate, for example, those using an additional program or intervention. Comparison groups must have been a no treatment or treatment-as-usual control for inclusion.

Table 3-1. Summary of inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Programs	Dating violence prevention programs	Sexual violence, general violence, peer violence
Participants	Adolescents (under 18)	Adult; targeted populations
Outcomes	Quantitative outcome measures relevant to knowledge about dating violence, attitudes towards dating violence, dating violence victimization or perpetration, or bystander behaviours	Qualitative measures; measure unrelated to dating violence
Study design	Design with a control or comparison group and/or a pre-test-post-test design	Post-test only designs
Sample size	At least 20 participants	Less than 20 participants
Language of publication	English	Non-English
Location of study	North America or Western Europe	Non-Western nations

3.2. Search strategy for identification of studies

3.2.1. Electronic searches

Using the set of search terms described below, 23 electronic databases were searched for potentially relevant articles (dates ranged from database inception to April 2019). These included:

- 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

3.2.2. Grey literature

Although electronic database searches can successfully identify a considerable number of relevant studies, focusing the search solely on these sources can result in a biased sample of studies. This is because bibliographic databases do not fully reflect the body of literature as a whole due to the exclusion of many other works. For a full, comprehensive search unpublished works must also be considered, such as theses and dissertations, conference presentations, technical reports, and independent projects. Sources that are not published in books or journal articles are referred to as the grey

literature and represent an importance part of a truly comprehensive search (Kugley et al., 2016).

Hand searching

Hand searching refers to a manual examination of additional sources, such as the websites of relevant organizations, curriculum-vitae of prominent researchers in the field, and non-database online search engines such as Google and Google Scholar. While this is a less systematic method of searching, it is sometimes useful in uncovering works not published in academic journals or indexed in bibliographic databases.

Google and Google Scholar were searched using more simplified search terms than those used in the electronic database searches (e.g., dating violence prevention program evaluation, effectiveness of dating violence prevention/relationship abuse prevention programs); the first 10 pages (100 hits) of each site were examined for potentially relevant studies. The curriculum-vitae of prominent researchers in the field were also searched for additional works not uncovered via the electronic databases. The specific authors were chosen based on the initial set of search results obtained using the electronic databases; authors were selected for curriculum-vitae searching if they had two or more studies identified as first author (see Appendix A.)

Additionally, a series of websites pertinent to the area of preventing violence against women were identified and searched for relevant studies. A Google search for “violence against women organization” was used to identify relevant organizations, in addition to national justice-focused organizations (e.g., Public Safety Canada, National Institute of Justice, Department of Justice). In total 16 organizations were identified and searched (full list in Appendix B). If the website had a search engine, the term “dating violence prevention program evaluation” was used to search the website and up to the first 10 pages of results were searched. The websites were also searched for any pages dedicated specifically to dating or relationship violence, and, when available, these pages were searched for relevant studies.

Backward searching

In addition to hand-searching, backward searching was used to identify applicable studies not uncovered by the electronic database searches. Backward searching involves examining the citations and reference lists of previously selected

studies. In an attempt to locate any theses, dissertations, reports, individual papers, or articles that had yet to be identified, the reference lists of all studies selected for inclusion were searched, as well as the reference lists and lists of included studies for any meta-analysis focused on dating violence prevention or relationship education programs (13 literature reviews and meta-analyses were identified and searched).

Table 3-2. Summary of search strategy

Search method	Number searched	Search terms
Electronic databases	23	Three detailed constructs (see section 3.3) combined using Boolean methods in the abstract field
Handsearching Google and Google Scholar Researcher CVs Organizations	 60 pages (600 hits) 20 16	 Simplified constructs (e.g. dating violence prevention program evaluation) used in available search engines. Curriculum vitae read to identify additional studies.
Backward searching	13 sources	Reference lists of relevant literature reviews and meta-analyses were searched for unidentified studies

3.3. Search terms

A set of three constructs was developed to conduct the systematic search and included: 1) dating violence, 2) prevention program, and 3) evaluation. As search engines use key terms to identify potentially relevant hits, it is critical that these constructs encompass a comprehensive and exhaustive list of terms. Key terms were first identified by consulting the relevant literature for common terms as well as synonyms and interchangeable phrases to ensure the search would identify any and all studies relating to dating violence prevention programs. Additionally, Boolean operators and wildcard marks were used to broaden the search. The constructs were developed using an iterative process of trial and error, with multiple rounds of testing. This was done to determine the correct combination of search terms that would locate as much relevant literature as possible, without identifying too many irrelevant studies that would inhibit the search process. The final iteration of the search strategy was applied to all 23 electronic databases and the key terms were searched in the Abstract field.

Construct 1: Dating violence

("dating violence" OR "relationship 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")

Construct 2: Prevention program

(program OR prevent* OR interven* OR campaign)

Construct 3: Evaluation

("eval*" OR "impact*" OR "outcome*" OR "assess*" OR "effect*")

3.4. Data collection and analysis

3.4.1. Selection of studies

Using the search terms listed above, the 23 electronic databases were searched and the available abstracts were reviewed. One reviewer read through the identified hits to determine which studies were potentially relevant to the current study and should be retrieved for further review. Once articles were deemed to meet initial eligibility criteria, a full list of studies was created to determine those that needed to be retrieved in full. When articles could not be located, the Simon Fraser University Inter-Library Loan (ILL) system was used.

Articles that were retrieved in full were examined by two reviewers to apply inclusion and exclusion criteria to determine inclusion for the study. If there were any discrepancies in reviewer decisions, the reasoning for the discrepancy was discussed until an agreement was reached regarding the inclusion or exclusion of a study. All included studies were then classified by age group (teen or college-aged)[3]. Studies selected for inclusion were coded by two independent reviewers. The interrater reliability

[3] Note that the search terms did not include an age criterion; age specification was applied during the coding process.

between the initial decisions for both retrieval and inclusion was assessed using Cohen's kappa. Cohen's kappa is a correlational measure of similarity across responses which accounts for the possibility that coders may guess on an inclusion decision due to uncertainty (McHugh, 2012).

3.4.2. Data extraction and management

Once studies were selected for inclusion, two reviewers independently coded each study on a series of 83 variables, using an Excel spreadsheet. These variables include program characteristics (e.g., program description, key program activities, main goals/purpose of program), intervention characteristics (e.g., number of sessions and contact hours, setting, who delivered the program material), study characteristics (e.g., research design, type of control group, if a pre-test was used), sample characteristics (e.g., total sample size at baseline, age range of participants, gender of participants), outcome measures (e.g., type of outcome measure, data presented, time of post-test), general study findings (e.g., overall findings), and treatment and control group outcomes (e.g., mean scores on measurement scales at pre-test and post-test). Using the same method described above, any disagreements were discussed until a consensus was reached.

3.4.3. Effect size calculations

One of the primary goals of meta-analysis is to synthesize research to draw an overall conclusion regarding the results of multiple studies (Card, 2011; Hedges, 2000; Lipsey & Wilson, 2001). To do so, however, the included studies must present data suitable for meaningful comparisons. Studies report findings using a variety of statistical forms (e.g., means, standard deviations, *t*-statistics, *F*-statistics, etc.) and research designs (e.g., RCTs, wait-list controls, single group pretest-post-tests). Due to the differing statistical methods and data reported, results across studies are not readily suitable for pooled analysis. As such, it is necessary to transform the various findings into the same metric: a standardized effect size (Card, 2011; Hedges, 2000; Morris & DeShon, 2002). The effect size statistic is the most widely used measure of treatment effect and is a standardized measure of the magnitude and/or direction of the relationship between two variables (Card, 2011; Lipsey & Wilson, 2001). These standardized and transformed effect sizes are computed for each included study, and

are then pooled to create a single, summary effect size to produce the overall meta-analytic conclusion. The current study calculated effect sizes based on five different types of reported outcome data: 1) pretest and post-test means with standard deviations (2 groups); 2) beta coefficient with standard deviations (2 groups); 3) *F*-test with unequal post-test sample sizes (2 groups); 4) percentage of participants who responded yes to the outcome (victimization or perpetration; 2 groups); and 5) pretest and post-test means with standard deviations and *t*-statistics or *F*-statistics (single group).

Combining two-group and single group research designs

In the field of meta-analysis, whether it is appropriate to pool effect sizes from single group pretest-post-test studies together with those from two-group designs is a controversial issue with no clear consensus. Many researchers argue that single group designs should not be included in meta-analysis for several different reasons. First, they suggest that scores on baseline and post-test measures are not independent of one another and the proper correlation between pre- and post-test scores is not accounted for when calculating a standardized mean difference (Cuijpers, Weitz, Cristea, & Twisk, 2017). By using incorrect correlations, the treatment effect could be considerably misestimated. Single group designs also lack the ability to disentangle the treatment effect from the effect of other factors (Carlson & Schmidt, 1999; Cuijpers, et al., 2017). The number of potential confounding variables can make the measures of treatment effect less reliable. Additionally, single group designs have been found to overestimate the magnitude of treatment effect; thus, including them in a meta-analysis could upwardly bias the pooled estimate (Lipsey & Wilson, 1993).

However, single group designs are often an important component of the research field, particularly in program and training evaluations. Field settings often come with situational and logistical constraints that prevent the use of a control group (Carlson & Schmidt, 1999). This is especially true in school settings, where issues regarding scheduling, access to students, and group assignment are common. In many cases, single group designs are the most feasible and/or the only option (Carlson & Schmidt, 1999). By excluding single group designs in meta-analyses, a considerable portion of the available research may not be considered in the overall conclusion of treatment impact. Though including single group designs may upwardly bias the pooled estimate, excluding them may also bias the estimate by not providing a full overview of the

treatment effect of interest (Card, 2011; Carlson & Schmidt, 1999). As well, excluding numerous single group studies may impact the ability for meta-analysts to conduct moderator analyses by limiting the number of studies included in the analysis overall (Carlson & Schmidt, 1999). Dating violence prevention programs, while generally similar in purpose and goals, often involve substantial differences in curriculum presented, teaching methods and activities, methods of delivery, length, and so forth. Thus, identifying factors that could moderate the treatment effect is critical for arriving at a robust conclusion regarding the effectiveness of such programs (Card; 2011; Carlson & Schmidt, 1999; Lipsey & Wilson, 2001). The exclusion of single group designs may run the risk of a resulting sample that is too small to enable the investigation of moderators.

While there is much criticism regarding the inclusion of single group pretest-post-test designs, some also argue that with certain adjustments, effect sizes from single group and two group designs can be appropriately pooled. Morris and DeShon (2002) state that for studies to be compared, they must provide estimates of the same treatment effect and must be computed in the same metric. Measuring the same population parameter is a critical element of meta-analysis; without equivalence across studies, the pooled estimate will not be a useful measure. Sometimes it may be reasonable to assume that the parameters between studies are equivalent or can be transformed to reach equivalence (Morris & DeShon, 2002). Part of this decision involves assessing the quality of the treatment estimate; while two-group designs inherently control for certain bias elements, single group studies do not. Thus, researchers must consider the level of control of potential bias and other factors that could impact the reliability of the treatment estimate in a given single group study (Borenstein & Hedges, 2019; Morris & DeShon, 2002). The decision of whether a study is appropriately measuring the treatment effect and if it should be included or excluded from analysis should be made on a case-by-case basis (Borenstein & Hedges, 2019; Morris & DeShon, 2002).

Additionally, Morris and DeShon (2002) state that effect sizes should be computed in the same metric, as differences in measures could lead to artificial differences in effect estimates. Meta-analysis uses a standardized effect size (e.g., standardized mean difference) to ensure the same metric. However, 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 (Morris & DeShon,

2002). Morris and DeShon (2002) suggest transforming one of the two types of effect sizes to correct for disparities in the metric and scaling. For example, if the majority of included studies are independent groups designs, this would require transforming the single group effect sizes into the raw score metric of the independent group effect size. Following these suggestions, the single group evaluations in the current study were assessed to determine if the treatment effect measured was comparable to that in the two-group studies. Those deemed comparable were included in the analysis after being transformed into a raw score metric (as per Morris & DeShon, 2002)[4].

Effect sizes for studies using a two-group design

The present study calculated effect sizes based on a total of four different types of data/research designs that used a two-group design: 1) pre and post means with standard deviations; 2) beta coefficient with standard deviations; 3) *F*-test with unequal post-test sample sizes; and 4) percentage of participants who responded yes to the outcome (victimization or perpetration).

1. pre and post means with standard deviations

The majority of the studies included in the current sample reported outcomes using continuous data (e.g., average score on a scale, number of questions answered correctly) and presented means and standard deviations (n = 16). For these studies, an effect size using Cohen's *d* was calculated. The basic standardized mean difference is calculated as the mean of the treatment group (M_T) minus the mean of the control group (M_C) divided by the pooled standard deviation (SD_{pooled}):

$$Cohen's\ d = \frac{M_T - M_C}{SD_{pooled}} \quad (1)$$

Where the pooled standard deviation is calculated by:

$$SD_{pooled} = \sqrt{\frac{(n_T - 1)SD_T^2 + (n_C - 1)SD_C^2}{(n_T - 1) + (n_C - 1)}} \quad (2)$$

[4] Sensitivity testing was performed to compare the transformed and non-transformed single group pretest-post-test effect sizes. Fixed and random effects models were conducted for all five of the primary outcomes and no substantive differences were found.

In which SD_T^2 is the standard deviation of the treatment group and SD_C^2 is the standard deviation of the control group.

The standard error is calculated by:

$$SE_{Cohen's\ d} = \sqrt{\frac{n_T + n_C}{n_T n_C} + \frac{d^2}{2(n_T + n_C)}} \quad (3)$$

To correct for possible small sample bias, Hedges' correction was applied to the Cohen's d equation, resulting in Hedges' *g*. Hedges' *g* was calculated for all relevant studies, as per the below equation:

$$Hedges'g = \left[1 - \frac{3}{4n - 9}\right]\left[\frac{M_T - M_C}{SD_{pooled}}\right] \quad (4)$$

Where $n = n_T + n_c$, n_T and n_T are sample sizes for the treatment and control groups, respectively, and SD_{pooled} is the pooled estimate of the standard deviation of the treatment and control groups

However, all (n = 16) of the included studies followed the pretest-post-test control group design, allowing for additional information to be factored into the effect sizes. The formula above was adapted to factor in the baseline data obtained at pretest. This allows for a more precise estimate; i.e., if there are group differences at pretest it could lead to a biased estimate if only the post-test data is used to calculate the effect size. By including the pretest data, we can compare the mean change from pretest to post-test in both the treatment and control groups.

$$ES_{SMD} = \left[1 - \frac{3}{4(n_T + n_C - 2)}\right]\left[\frac{(M_{T,post} - M_{T,pre})(M_{C,post} - M_{C,pre})}{SD_{pre}}\right] \quad (5)$$

Using a small samples bias corrector, $M_{t,\ post}$ is the post-test mean of the treatment group, $M_{t,\ pre}$ is the pretest mean for the treatment group, $M_{c,\ post}$ is the post-test mean for the control group, $M_{c,\ pre}$ is the pretest mean for the control group, and SD_{pre} is the pooled standard deviation at pretest (Morris, 2008), calculated as follows:

$$SD_{pre} = \sqrt{\frac{(n_T - 1)SD_{T,pre}^2 + (n_C - 1)SD_{C,pre}^2}{n_T + n_C - 2}} \quad (6)$$

And the standard error is calculated as:

$$SE_{SMD} = \sqrt{\frac{n_T + n_C}{n_T n_C} + \frac{(ES_{SMD})^2}{2(n_T + n_C)}} \quad (7$$

2. *beta coefficient with standard deviations*

It is common for studies to fail to report the means and/or standard deviations; in these cases, it is possible to compute a standardized mean difference using other available inferential statistics (Lipsey & Wilson, 2001).

One study (Joppa, 2016) in the current set presented an unstandardized beta coefficient, standard deviation of the dependent variable, and sample sizes. A small sample bias corrector was included and the effect size was calculated as follows, where the SD_{pooled} *(equation 6)* is calculated using pre-test *n's*:

$$ES = \left[1 - \frac{3}{4n - 9}\right]\left[\frac{\beta}{SD_{pooled}}\right] \quad (8)$$

Where the standard error is calculated by:

$$SE = \sqrt{\frac{n_T + n_C}{n_T n_C} + \frac{(ES)^2}{2(n_T + n_C)}} \quad (9)$$

3. F-test with unequal post-test sample sizes

One study (Chamberland, 2014) in the current set presented an *F*-test with unequal sample group sizes. The effect sizes for this study were calculated as follows:

$$ES = \sqrt{\frac{F(n_T + n_C)}{n_T n_C}} \quad (10)$$

Where *F* is the *F*-statistic, n_T is the post-test sample size for the treatment group and n_C is the post-test sample size for the control group. Standard error is calculated as per equation 9.

4. *percentage of participants who responded yes to the outcome (victimization or perpetration)*

Eight of the included studies presented dichotomous outcome measures (often related to violence perpetration and victimization outcomes). This was typically in the form of percentages or raw numbers representing how many participants experienced or perpetrated incidents of dating violence. Effect sizes for these studies are computed as odds ratios, which equate to comparing those who experienced violence to those who did not; the odds refer to the chance of something happening compared to the chance it will not happen (Lipsey & Wilson, 2001). However, the odds ratio measure of effect size is not numerically compatible with the standardized mean difference; a common method used for combining effect sizes from studies that use continuous and dichotomous data is to apply a transformation to adjust for the dichotomous data (Lipsey & Wilson, 2001). The current analysis calculated log odds ratios and used the Cox transformation, which involves dividing the log odds ratio by 1.65 (Sanchez-Meca, Marın-Martinez, & Chacon-Moscoso, 2003):

$$ES_{cox} = \frac{L(OR)}{1.65} \tag{11}$$

Where the standard error is calculated as:

$$SE_{cox} = \sqrt{.367\left[\left(\frac{1}{a}\right) + \left(\frac{1}{c}\right) + \left(\frac{1}{b}\right) + \left(\frac{1}{d}\right)\right]} \tag{12}$$

In which *a* is the number of treatment participants who reported no perpetration/victimization, *c* is the number from control who reported no perpetration/victimization, *b* is the number of treatment participants who reported yes to perpetration/victimization, and *d* is the number from the control group who reported yes to perpetration/victimization.

Given that these eight studies included both pretest and post-test data, this formula was also adapted to take this information into account. The pretest-adjusted formula used is as follows:

$$ES_{cox} = \left[\frac{L_{OR,post}}{1.65}\right] - \left[\frac{L_{OR,pre}}{1.65}\right] \quad (13)$$

With the standard error calculated as:

$$SE_{cox} = \sqrt{.367\left[\left(\frac{1}{a_{pre}}\right) + \left(\frac{1}{c_{pre}}\right) + \left(\frac{1}{b_{pre}}\right) + \left(\frac{1}{d_{pre}}\right)\right]} \quad (14)$$

In which a_{pre} and c_{pre} are the successful outcome frequencies (i.e., number of treatment and control participants who reported no perpetration/victimization at pretest, respectively), and b_{pre} and d_{pre} are the number of non-successful outcome frequencies (i.e., treatment and control participants who reported yes to perpetration/victimization at pretest).

Effect sizes for studies using a single group pretest-post-test design

5. *pre and post-test means with standard deviations and t-statistics or F-statistics (1-group).*

Eleven studies used a single group pretest-post-test design and presented either a *t*-statistic (n = 7) or an *F*-statistic (n = 4). As per Morris and DeShon (2002), an adapted version of Cohen's *d* was calculated using the pretest and post-test means and standard deviations (rather than means for treatment and control groups):

$$ES_{sgpp} = \frac{M_{post} - M_{pre}}{SD_d} \quad (15)$$

Where M_{post} is the post-test mean, M_{pre} is the pretest mean and SD_d is the standard deviation of the difference scores. SD_d is calculated as:

$$SD_d^2 = \frac{n\left(M_{post} - M_{pre}\right)^2}{t_{sgpp}^2} \quad (16)$$

Where M_{post} is the post-test mean, M_{pre} is the pretest mean and t^2_{sgpp} is the single group pre-test post-test test statistic (when a *t*-test is presented). The following equation was used to calculate SD_d when a *F*-test was presented:

$$SD_d^2 = \frac{n\left(M_{post} - M_{pre}\right)^2}{F_{sgpp}} \tag{17}$$

The standard error was calculated using the following formula (for both *t*- and *F*-tests):

$$SE_{sgpp} = \sqrt{\frac{2(1-r)}{n} + \frac{ES_{sgpp}^2}{2n}} \tag{18}$$

These effect sizes were then transformed from a change score metric to a raw score metric to be commensurable with the effect sizes produced from two-group designs (pre-test post-test with control; Morris & DeShon, 2002):

$$ES_{ppwc} = ES_{sgpp}\sqrt{2(1-r)} \tag{19}$$

Where ES_{sgpp} is the original repeated measures effect size and r is the pretest-post-test correlation:

$$r = \frac{SD_d^2}{2\left(SD_p^2\right)} \tag{20}$$

Where SD_d is the standard deviation of the difference scores and SD_p is the pooled standard deviation.

Table 3-3. Summary of effect size calculations

Type of data used to calculate effect sizes	Number of studies	Number of effect sizes
Two-group designs		
Pre and post means with standard deviations	16	34
Beta coefficient with standard deviations	1	2
F-test with unequal post-test sample sizes	1	2
Percentage of participants reporting violence	8	15
Single group designs		
Pre and post-test means with standard deviations and *t*-statistics or *F*-statistics	11	17

Cluster adjustments

Many of the studies included did not assign individual students to the treatment or comparison groups; instead, conditions were assigned based on various groups such as middle schools, high schools, or school classrooms. Given that participants in these designs were nested within clusters, it was necessary to cluster-adjust the effect sizes and standard errors to account for the nesting (Hedges, 2007). These adjustments were done using an estimate of the intra-cluster correlation coefficient (ICC or rho), a measure of the proportional variance attributable to group differences. Of the 37 studies included, only six reported an exact ICC or a range of ICCs. For the remaining studies, we turned to the literature to determine an appropriate estimate of the ICC for each cluster type. When examining the literature regarding proper ICC estimates for the various cluster levels, it became apparent that guidelines are lacking. Very little information on suggested ICC values for relevant cluster and outcome types was available, with the What Works Clearinghouse (WWC, 2017) being one of the few options. However, the WWC guidelines are quite conservative, which is worth consideration as the ICC used in the cluster-adjusting calculations can have a considerable impact on the resulting standard errors. This in turn can impact how a study is weighted in the meta-analysis, which has the possibility to substantially change the overall pooled effect.

Table 3-4. Summary of cluster-adjustments

Cluster Adjustment	Number of studies	Number of effect sizes
Cluster adjusted	21	45
Non-cluster adjusted	16	25

School (middle or high school)

Of the studies with clusters at the school level (n=13), six reported ICCs. The ICCs ranged from .001 to .04, and were used in the cluster-adjustment calculations for their corresponding study (Levesque et al., 2016; Miller et al., 2012; Peskin et al., 2014; Reidy, Holland, Cortina, Ball, & Rosenbluth, 2017; Sanchez-Jiminez, Munoz-Fernandez, & Ortega-Rivera, 2018, and Wolfe et al., 2009). These reported ICC values were used in conjunction with other existing literature to determine an overall ICC estimate for school-level clusters for the remaining seven studies that did not report specific estimates. For example, Murray and Blitstein (2003) examined ICC estimates across six studies with 342 ICCs and found an average ICC of .026 for behavioural outcomes and an average ICC of .028 for attitudinal outcomes. Additionally, Spence, Sheffield, and Donovan (2003) reported an ICC of .0225 for school-level outcomes. Using the findings regarding ICC values from these two studies, in addition to the reported ICC values from the studies within the current sample, we concluded that an ICC estimate of .026 for behavioural outcomes and an ICC estimate of .028 for attitudinal outcomes was appropriate for the school-level cluster adjustments.

Classroom

For classroom-level clustering (n = 8), no studies within the present sample reported exact ICC values. As such, we relied on existing literature to determine an appropriate ICC estimate for classroom-level clusters and outcomes. The WWC (2017) provides general guidelines regarding educational types of outcomes; for behavioural and attitudinal outcomes in an educational setting, the WWC recommends an ICC of .10. When examining the literature involving similar behavioural and attitudinal outcomes as those in the present study (e.g., bullying behaviours, attitudes towards bullying, aggression perpetration and victimization), the reported ICCs are varied, with a range of .02 to .40 (Brackett, Reyes, Rivers, Elbertson, & Salovey, 2011; Di Stasio, Savage, & Burgos, 2016; Fonagy, et al., 2009; Isaacs, Voeten, & Salmivalli, 2013; Meilstrup, et al.,

2015; Oberle & Schonert-Reichl, 2016; Saarento, Boulton, & Salmivalli, 2015; Salmivalli, Voeten, & Poskiparta, 2011; Thornberg, Wanstrom, Hong, & Espelage, 2017). As the literature included a large range of ICC values, we conducted cluster adjustments using both the WWC recommendation of .10 and a smaller estimate of .05 (based on the existing literature consulted). Using both sets of cluster-adjusted effect sizes, we conducted sensitivity analyses for each of the five outcome measures (dating violence knowledge; attitudes towards dating violence; incidents of dating violence perpetration; experiences of dating violence victimization; and bystander behaviours) to determine if the two ICC estimates resulted in significant differences for overall pooled conclusions using both fixed effects and random effects models. Both ICC estimates resulted in very similar results in both types of models for all outcomes. As such, the WWC recommendation of .10 was selected as the ICC for the final analyses.

Studies with multi-level clustering

Several of the included studies employed multiple levels of clustering (e.g., by school as well as by classroom), and as such required a decision as to which level of clustering to adjust for. According to the U.S. Department of Health and Human Services (2013), if an intermediate level of clustering is not randomly assigned, but the primary level is, then it is only necessary to adjust for clustering at the primary level. For example, if the primary cluster level of school is randomly assigned to a condition, and the school has multiple health classrooms that are not randomly assigned, then only cluster adjustments at the school-level are required.

The current analysis included seven studies that had more than one level of clustering. Of the seven, four were randomly assigned at the school level; based on the information above, these four studies were cluster-adjusted using the school-level ICC estimates. One study (Banyard et al., 2019) did not have any random assignment and was cluster-adjusted using the school-level ICC estimates as these estimates were more conservative than the classroom-level ICC options. One study (Cissner & Ayoub, 2014) randomly assigned 75% of the sample at the classroom level and 25% at the individual level. As the classroom-level assignment accounted for the majority of the sample, this study was cluster-adjusted using classroom-level ICC estimates. The final study of the three was randomized using “tracks”, which were determined to equate to subgroups

within a school that operated on different schedules (Jaycox et al., 2006). As such, this study was cluster-adjusted using the classroom-level ICC estimates.

3.4.4. Outcome measures

The present study included a series of five outcome measures: 1) knowledge of dating violence, 2) attitudes and beliefs towards dating violence, 3) perpetration of dating violence behaviours, 4) experiences of dating violence victimization, and 5) bystander behaviours and intentions.

Knowledge of dating violence

Measures of dating violence knowledge included those focused on the ability to recognize and identify types of dating violence/dating violence behaviours or facts about dating violence. Studies typically produced their own scale of measurement specific to the information presented in the program or measures used in previous evaluations (however, the measures were not necessarily validated). Knowledge was primarily measured as the number of correct answers on a scale or series of questions. Questions included those such as: “Violence in relationships happens more often to teenagers than to adults” (Connolly et al., 2015) or “The first stage of a violent act is when someone says or does something that makes you uncomfortable. This is called (a) Desensitization. (b) Intrusion. (c) Eye of the Beholder. (d) Isolation” (McLeod, Jones & Cramer, 2015).

Attitudes and beliefs towards dating violence

This measure included those assessing attitudes or opinions of dating violence and aggression (e.g., beliefs about if/when violence is appropriate) or acceptance of rape myths. Specific measures varied for this outcome, with some studies using self-made measures, while others used pre-existing scales. Two of the commonly used scales are the Acceptance of Couple Violence scale (Foshee, Fothergill & Stuart, 1998) and the Attitudes Towards Dating Violence Scale (Price & Byers, 1999). Example items from the latter scale include: “It is OK for a guy to badmouth his girlfriend”; “It is OK for a girl to slap her boyfriend if he deserves it”; and “When a guy pays on a date, it is OK for him to pressure his girlfriend for sex” (Connolly et al., 2015).

Perpetration of dating violence behaviours

Measures of dating violence perpetration included self-reported aggressive behaviours in a dating relationship (e.g., hitting, slapping, name-calling). Incidents of dating violence perpetration were most often measured using either the Conflict in Adolescent Dating Relationships Inventory (Wolfe et al., 2001) or the Conflict Tactics Scale (or a modified version thereof; Straus, 1979). The Conflict Tactic Scale requires respondents to rate how many times they had perpetrated an act against their boyfriend or girlfriend (range = 0 – 20+ times in the last two months); example behaviours include: “yelled at him or her” or “slapped him or her” (Gardner et al., 2015).

Experiences of dating violence victimization

This measure included any outcome involving self-reported incidents of being victimized by a dating partner (e.g., being hit, slapped, called names). Victimization experiences were most commonly measured using the Conflict in Adolescent Dating Relationships Inventory to rate how often respondents were victimized (Wolfe et al., 2001). Example items include “During a conflict or argument I was kicked, hit, or punched by my boy/girlfriend”; “He or she tried to turn my friends against me”; and “He or she made fun of me in front of others” (Connolly et al., 2015).

Bystander behaviours and intentions

Last, the bystander measures refer to self-reported instances of engaging in bystander behaviour (e.g., speaking up when rape myths are perpetuated, telling a peer to stop hitting their partner) or measures of intentions to engage in bystander behaviour (e.g., “if your friend called their partner a rude name, how likely is it that you would say something?”). The most common measure used by studies for this outcome include the Bystander Behaviours Scale (Banyard, Moynihan, Cares, & Warner, 2014). Example items include yes/no questions such as: “I saw a friend in a heated argument. I asked if everything was okay” and “I confronted a friend who made excuses for the abusive behaviors of others” (Jouriles, McDonald, Rosenfield, & Sargent, 2019).

Table 3-5. Summary of outcome measures

Outcome measure	Examples of measures used	Example items from measures
Knowledge of dating violence	Scales made by researcher	"Violence in relationships happens more often to teenagers than to adults" "The first stage of a violent act is when someone says or does something that makes you uncomfortable. This is called (a) Desensitization. (b) Intrusion. (c) Eye of the Beholder. (d) Isolation"
Attitudes and beliefs towards dating violence	Acceptance of Couple Violence scale (Foshee, Fothergill & Stuart, 1998) Attitudes Towards Dating Violence Scale (Price & Byers, 1999).	"It is OK for a guy to badmouth his girlfriend" "It is OK for a girl to slap her boyfriend if he deserves it"
Perpetration of dating violence	Conflict in Adolescent Dating Relationships Inventory (Wolfe et al., 2001) Conflict Tactics Scale (Straus, 1979)	Respondent asked to rate how many times they had perpetrated an act against their boyfriend or girlfriend (range = 0 – 20+ times in the last two months)
Experiences of dating violence victimization	Conflict in Adolescent Dating Relationships Inventory (Wolfe et al., 2001)	"During a conflict or argument I was kicked, hit, or punched by my boy/girlfriend" "He or she made fun of me in front of others"
Bystander intentions or behaviours	Bystander Behaviours Scale (Banyard, Moynihan, Cares, & Warner, 2014)	"I saw a friend in a heated argument. I asked if everything was okay" "I confronted a friend who made excuses for the abusive behaviors of others"

3.4.5. Decision rules

Another important factor to consider in the selection of studies is that of independence of effect sizes. 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). Nonindependence can arise as a concern when multiple studies present results based on the same population or sample, or when studies present effect sizes from multiple measures (Card, 2011). There are different ways for handling situations of nonindependence, such as averaging across effect sizes or study samples/subgroups or selecting individual effect sizes based on decision rules that align with the inclusion and exclusion criteria (Card, 2011). The latter option was chosen for the present study and the decision rules applied are as follows:

1. ***Multiple studies based on the same population or sample:*** If more than one study reported results based on the same population, the first published report was chosen. Often the earlier report presented more relevant outcome measures (additional reports typically presented additional follow-up results or subgroup analyses). However, if outcomes differed between studies, then the relevant independent outcomes from both were selected.
2. ***Multiple reports of the same study:*** If multiple documents presented data from the same study, the one with the most detailed information was chosen. For example, if a journal article and a technical report of the same study were identified, and the technical report included more detailed results, the report was chosen over the journal article.
3. ***Reporting on separate subgroups from the same study, in the same document:*** If a study separated their sample into independent subgroups (e.g., by gender), results were combined across subgroups to create a single effect size. In other words, rather than treating each subgroup as an independent effect size, a composite effect was used (see Borenstein, Hedges, Higgins, & Rothstein, 2009).
4. ***Multiple studies in the same document:*** If a document reported on more than one experiment in the same report, both were coded as independent studies, unless there was any overlap between samples.
5. ***Multiple treatment groups:*** if a study included multiple treatment groups and only one control group, the treatment group that was the most relevant and the most comparable to other included treatment groups/programs in the overall set of studies was selected for inclusion.
6. ***Multiple post-tests:*** 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 among all included studies. The majority of the studies in the set administered an immediate post-test, and while many studies also included an additional follow-up measurement, the timing of the follow-ups varied across studies (ranging from one month to over two years). Thus, the immediate post-test was the most consistent measure across studies and was chosen as the point of measurement for analysis, *except* when the statistical information required for the calculation of an effect size was only available for a later follow-up time point. In these cases, the point of follow-up chosen was that which corresponded to the analysis reported.
7. ***Multiple outcomes in a single study:*** Some studies reported on multiple outcomes within the same study; in these situations the most relevant and commensurate outcome was selected. For example, if a study presented results for knowledge of both "smart dating strategies" and "unhealthy relationships" (Adler-Baeder, 2007), "unhealthy relationships" was chosen as it was deemed more relevant to the overall topic of dating violence knowledge and more commensurate with other studies in the overall set.
8. ***Multiple measures of dating violence perpetration/victimization:*** when a study reported multiple measures of violence perpetration/victimization using the same sample (e.g., physical, emotional, sexual violence), the options were prioritized in the order of emotional, physical, sexual. Emotional violence was prioritized due to the age of the target population (i.e., emotional/ psychological violence is highly reported among youth (Leen et al., 2012; Taylor & Mumford, 2016)). Some studies reported a measure of "any" violence perpetration/victimization; in these cases, the outcome was classified into one of the three types of violence based on the items that

comprised the measure. For example, Ball (2012) included the full list of items comprising the "any violence" measures and, of the 10 items, 8 were related to emotional/psychological violence and the effect size was thus coded as an emotional violence outcome.

Dealing with missing data

Due to inconsistencies in the reporting of program characteristics, study design, and study outcomes, there were often instances of missing data. When appropriate, the authors were contacted with a request for information or clarification regarding aspects of the study. Unfortunately, this endeavor proved unsuccessful in obtaining any of the missing data. As such, when possible the missing data were inferred or calculated using available data. In particular, 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. When available, 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 pretest sample sizes. For example, Adler-Baeder et al. (2007) reported treatment and control sample sizes at pre-test, but did not provide post-test sample sizes after attrition and casewise deletion.[5]

Another data category that had considerable reporting inconsistency was that of program components. During the data extraction process, several potential program components or characteristics were coded for (e.g., whether the curriculum addressed gender roles, if role-play activities were included, if group discussion was involved). However, there was substantial inconsistency in reporting whether certain program components were included in the curriculum; in other words, it was sometimes difficult to determine if a program truly did not include a component or if the study's description of the program curriculum simply did not mention it. For these instances, it was often possible to infer the information based on other information provided (e.g., it was inferred

[5] I.e., the treatment sample was 235 and control sample was 105, for a total of 340 participants. For the verbal aggression effect size, the analysis sample was 299, which meant 41 participants were dropped from the analysis. Since the treatment group accounted for 69% of the pretest sample and the control group accounted for 31%, we assumed that 69% of the dropped participants were from the treatment group and 31% of the dropped participants were from the control group.

that group discussion was involved if the program was delivered in-person in a small group setting). If coders were approximately 85% confident or more that the component was *not* included in the program, that study was coded as a “no”; if they were less than 85% confident, the study was coded as “missing”. The same decision rule applied to “yes” decisions; if coders were at least 85% confident the component *was* included in the program, the study was coded as a “yes”, if not, it was coded as “missing”. This was to minimize the amount of missing data to allow for analyses examining these program characteristics.

3.4.6. Data synthesis (meta-analysis)

In meta-analysis, the included effects sizes are pooled together to create a single mean effect size representing the overall treatment effect (Lipsey & Wilson, 2001). However, studies differ in methodological quality, which impacts the levels of precision in the estimates of treatment effect. For example, studies with larger samples are often associated with smaller variances, resulting in a more precise estimate, while smaller samples are subject to greater sampling error and less precise estimates (Konstantopoulos & Hedges, 2019; Lipsey & Wilson, 2001). This difference would be problematic if the studies with differing levels of precision were simply averaged, as both would have an equal impact on the pooled result, despite the disparities in sampling error. To address this issue in meta-analysis, each study is weighted by its inverse variance weight, which is calculated using the standard error of the effect size (Lipsey & Wilson, 2001). This technique allows for more precise study estimates to contribute greater weight to the pooled effect, increasing the accuracy of the mean estimate.

Fixed effects model

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 exact information used in calculating the weights differs based on the assumed source of variability between studies (Card, 2011; Egger & Smith, 2001; Lipsey & Wilson, 2001). A fixed effects model assumes the between-subject variability present is the result of sampling error and occurs only by chance (Lipsey & Wilson, 2001). In this model, all other aspects of the studies are assumed to be similar such that there are no other factors contributing to the variability other than chance. Based on this assumption,

studies are weighted simply by their inverse variance, which, as described above, allows for the maximization of estimate precision. Less precise studies are weighted less heavily than more precise studies.

Random effects model

An alternative to the fixed effects model is the random effects model. This model assumes that between-study heterogeneity is significant and the variability is due to factors other than random subject-level sampling error (Lipsey & Wilson, 2001). While the random effects model also weights each study by its inverse variance, the model also considers other factors that could be contributing to the variability in the calculation of the weights; sources of both subject-level and study-level error are factored in (Card, 2011; Lipsey & Wilson, 2001). This model typically results in more conservative estimates, with smaller, more similar weights across studies compared to those of a fixed effects model (i.e., in comparison to a fixed effects model, larger studies are weighted less heavily, while smaller studies are weighted more heavily; Card, 2011).

Given the nature of the sample of studies in the present analysis, a significant level of heterogeneity is expected. As the specific programs being evaluated in each study differ from one another, in addition to a multitude of other study differences, the variability among studies is likely due to factors other than random subject-level sampling error alone. For this reason, random effects models were chosen for the primary analysis.

3.4.7. Sensitivity analysis

Publication bias

In meta-analysis, publication bias is a particularly problematic issue. The aim of meta-analysis is to synthesize research into a single conclusion; however, only studies that are available and accessible can be included within the synthesis. Often the available studies are the product of publication bias: the notion that studies that produce significant, positive results are more likely to be published than are studies with nonsignificant findings (Lipsey & Wilson, 2001; Sterne & Harbord, 2004). Similar to this concept is that of small study effects; studies with small samples are more likely to overestimate the treatment effect as studies with small samples typically have larger

sampling errors and require a larger treatment effect to result in a significant finding (Egger, Juni, Bartlett, Holenstein, & Sterne, 2003; Sterne & Harbord, 2004).

Publication bias was assessed using two methods. The first was by using a funnel plot, which is a graphical representation of the estimated effect size plotted against the standard error (Sterne & Harbord, 2004). This scatterplot provides a visual estimate of the levels of precision among the included studies; if there is no bias present, the plots will resemble the shape of an inverted funnel. Larger, more precise estimates will lie closer together at the upward point of the funnel, while smaller studies will scatter along the bottom of the plot. The overall funnel will fall within the constraints of pseudo 95% confidence intervals, which illustrate the expected distribution in the absence of heterogeneity; they are referred to as 'pseudo 95% confidence intervals' because they are not strict limits (Sterne & Harbord, 2004). If publication bias is present, we would expect to see a void in the lower, right-hand section of the graph (Sterne & Harbord, 2004). This is due to the lack of small studies without statistically significant effects. It is important to note that funnel plots are not proof of bias, but if asymmetry is found, it is worthwhile to consider the possible sources (Sterne & Harbord, 2004).

Additionally, Egger's test of small study effects was performed as it can provide information on whether small study effects are present, and, by extension, if publication bias is present (Steichen, 1998). This test regresses the standardized effect sizes against their precision estimates; if publication bias is present, the small sample studies will differ significantly from the large sample studies, with the intercept deviating significantly from zero. Therefore, a greater deviation from zero represents greater asymmetry (Steichen, 1998).

Influence analysis

Again, as the goal of meta-analysis is to synthesize research results, it is important to ensure that the included studies are comparable and the pooled effect is not being distorted. One issue to consider is the effect of outliers; in other words, outlier effects from studies that differ considerably from others in the included sample and may be having a disproportionate influence on the pooled estimate (Lipsey & Wilson, 2001). To assess potential outlier effects, influence analysis can determine whether any individual study is having a substantial impact on the pooled effect (Tobias, 1999). With this approach, each study is removed one at a time to determine the resulting pooled

estimate when that study is omitted from the analysis. If the pooled estimate changes drastically or changes the significance level of the pooled effect, the effect can be deemed sensitive and identified outliers can be adjusted to less extreme values or omitted from the analysis altogether (Lipsey & Wilson, 2001).

3.4.8. Assessment of heterogeneity

When combining multiple effect sizes to compute a summative effect size, it is important to understand whether the included effect sizes are measuring the same population effect (Lipsey & Wilson, 2001). If the effect sizes are measuring the same effect and are thus homogenous, then the differences among the means should be the result of subject-level sampling error alone (Lipsey & Wilson, 2001). However, if the effect sizes are heterogeneous, this suggests the studies are not measuring the same population effect and that there may be other sources of between-study differences that impact the estimates.

Two statistics were used to assess heterogeneity in the current analysis: the *Q*-statistic and the I^2 statistic. The *Q*-statistic represents a measure of between-study heterogeneity. If *Q* is significant, we reject the null hypothesis of homogeneity and assume there is a level of heterogeneity between studies that is due to more than just sampling error (Card, 2011; Lipsey & Wilson, 2001). In comparison, the I^2 statistic is a measure of the magnitude of heterogeneity and how much heterogeneity can be attributed to sources other than sampling error (Card, 2011). If I^2 is greater than 50%, it suggests a high level of variability which is attributable to heterogeneity caused by factors other than sampling error or chance (Card, 2011).

Subgroup analysis

When a sample is significantly heterogeneous and has a high I^2, it is important to investigate the potential sources of heterogeneity impacting the pooled estimate (Card, 2011; Lipsey & Wilson, 2001). A variety of factors could influence levels of between-study heterogeneity, such as program characteristics (e.g., whether the program addresses gender roles and stereotypes, if the program incorporates group discussion or activities, length of program) or study characteristics (e.g., sample size, strength of research design). These factors can be investigated as sources of heterogeneity through subgroup (moderator) analysis.

One method of subgroup analysis is the analog to the ANOVA. This approach tests whether a categorical variable can explain the variability in effect sizes (Card, 2011; Lipsey & Wilson, 2001). The analog to the ANOVA separates the total variability (Q_T) into that which can be explained by the categorical variable (the between-group variation in the characteristic of interest; Q_b) and the remaining within-group variation(Q_w). More specifically, the three statistics needed to conduct subgroup analysis using the analog to the ANOVA method are defined as:

Q_T = Q-total; the overall heterogeneity produced by the full set of studies using either a fixed or random effects model

Q_w = Q-within (Q_{group1} + Q_{group2}); the summed Q-statistics for each of the two groups in the categorical analysis[6]

Q_b = Q-between (Q_T – Q_w); 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 (Lipsey & Wilson, 2001). In this case, we could conclude that the variable is important for moderating the results and that it is contributing to the variability beyond sampling error.

To assess heterogeneity, a series of nine dichotomous variables were selected as potential moderating variables. A preliminary list of possible variables was first selected based on existing literature. Program characteristics (e.g., the curriculum addressed gender roles and stereotypes, the program incorporated group discussion or activities) were selected based on current literature that suggests these components may impact the program effect (e.g., Foshee et al., 2004). Study characteristics (e.g., sample size, research design) were also selected based on literature that suggests they may be related to study outcomes (e.g., Lipsey & Wilson). Additionally, characteristics were selected based on overall availability in the dataset. Those variables that were reliably reported on or could be consistently inferred from the information provided were chosen; variables that were available in the current dataset were then examined to determine suitability for subgroup analyses. Correlations and chi-square analyses were

[6] For example, if the categorical variable was sample size, Q_{group1} may refer to the *Q*-statistic produced by samples smaller than 300 participants, while Q_{group2} may refer to the *Q*-statistic produced by samples equal to or larger than 300 participants.

conducted to assess collinearity; variables that were highly correlated were not used for analysis. The nine moderating variables used in the current analysis are as follows:

Program characteristics:

1. Length of program (≤ 5 hours versus ≥ 6 hours)
2. Curriculum addressed gender roles and stereotypes (yes/no)
3. Curriculum incorporated bystander training (yes/no)
4. Level of school (middle school versus high school)

Study characteristics:

1. Treatment group sample size (< 300 participants versus ≥ 300 participants)
2. Research design (randomized control trial versus non-randomized design)
3. Publication year (prior to 2015 versus 2015 or later)
4. Time of post-test (immediately after program end versus 1+ months after program end)
5. Sample ethnicity (group was predominantly Caucasian or mixed ethnicity versus predominantly an ethnic minority)

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, incidents of perpetration and victimization of dating violence, and bystander behaviours). Also, subgroup analysis was conducted for only four of the five outcomes, due to the small number of effect sizes for the fifth outcome (bystander behaviours).

Chapter 4.

Results

4.1. Results of the systematic search

The initial search of the 23 electronic databases resulted in a total of 4,430 articles for review, while the search of the grey literature identified 962 additional hits for review (not including the potential sources from the curriculum-vitae of authors or reference lists of included articles and related meta-analyses). The abstracts of all hits were reviewed and 135[7] articles were identified as potentially meeting the inclusion criteria and were retrieved in full. The full inclusion and exclusion criteria were applied to the 135 articles by two reviewers, who selected 88 studies for inclusion. Interrater reliability was high for the retrieval agreement with a kappa of 0.7856 (z = 10.57, p < .001), and moderately high 0.537 (z = 6.22, p < .001) for the inclusion agreement. During both the retrieval and inclusion processes, if agreement was not met we erred on the side of caution and retrieved/included a study despite only one reviewer coding it as "yes". During the coding and extraction process, studies were categorized based on target population age (teen or college)[8]. Due to missing data, data that were inappropriate for calculating commensurable effect sizes, or overlapping samples, an additional 39 studies were excluded during the data extraction and effect size calculation processes[9]. This left a sample of 49 studies to be included in the final sample. However, as the current analyses were focused on programs targeting an adolescent population, the college programs were excluded, leaving a final sample of 37 independent studies which contributed 71 independent effect sizes.

[7] We initially identified 139 articles but were unable to retrieve 22; these were requested using the Inter-Library Loans system through the SFU Library. Of the 22, 18 were retrieved in full, leaving 4 unable to be located and resulting in a final 135 articles which were reviewed in full.

[8] The initial search was not limited to teen populations; the full search included all age groups, such as college-aged adults.

[9] In total, 14 studies were excluded due to missing data; 26 studies were excluded due to data inappropriate for calculating effect sizes; 11 studies were excluded due to irrelevant outcomes; 7 studies were excluded due to overlapping samples; 6 were excluded due to an inappropriate comparison group; 3 studies were excluded due to small sample size; and the remaining 19 studies were excluded due to other reasons (e.g., not being an evaluation, published in a non-Western country, not published in English, targeted a too specific population, or was a post-test only design).

4.1.1. Characteristics of the included studies

Overview of the included studies

From the 37 studies, 71 effect sizes were calculated and were dispersed across the five outcome measures: knowledge about dating violence (27%), attitudes towards dating violence (22%), incidents of dating violence perpetration (22%), incidents of dating violence victimization (16%), and bystander behaviours and intentions (13%). Most programs were implemented in North America (81% in the United States; 8% in Canada), while 11% were implemented in Europe.

Publication date and type

Regarding publication year, studies were nearly equally published before and after the year 2015; however, the majority of studies were published in the last nine years, with only 11 (30%) studies published prior to 2011. Additionally, the majority (81%) of the included studies were from peer-reviewed sources (journal articles), with only seven of the studies being non-peer-reviewed documents such as theses, dissertations, or technical reports.

Table 4-1. Publication characteristics of the included studies (n = 37)

Study Characteristics	*n* (%)
Publication Year (range 1997-2019)	
Pre-2015	19 (51.35)
2015-2019	18 (48.65)
Publication Type	
Journal	30 (81.08)
Reports	2 (5.41)
Dissertation/Thesis	5 (13.51)

Program components

Several program characteristics and components were coded for. These include year of program implementation, program location, number of sessions and number of hours, method of delivery, setting, facilitator(s), and elements of the curriculum (specifically, if the program addressed gender roles, included bystander training, focused on skill development, included role play, include group activities, or used group discussion).

All programs were implemented between the years 1994 and 2017 with a majority (59%) implemented in the year 2010 or later. Program length varied considerably in both the number of sessions and the total number of hours spent in the program. The number of sessions ranged from 1 to 25, while the number of hours ranged from 15 minutes to 28 hours. Approximately 60% of studies were between 2 and 10 hours long, with 2-10 sessions in total. A majority of the studies evaluated programs that were delivered in person, either in school or in a community setting (e.g., a community center), with 95% of studies involving an in-person program and 5% focused on an online program. Programs were delivered in several different settings, including classroom (73%), small group (not in school/class, e.g., in a community setting or afterschool; 16%), large group (e.g., a theatre performance; 3%), and online (5%), with one study classified as other (a one-on-one discussion in a school health center; 3%). Additionally, programs involved several different types of facilitators, with 24% of programs delivered by teachers, 27% by program facilitators, 29% by other professional staff (e.g., counsellors, social workers), 8% by peers, and 5% via an online platform.

Table 4-2. Program characteristics of the included studies (n = 37)

Characteristics	*n* (%)
Program year	
Pre-2010	15 (40.54)
2010-2017	22 (59.46)
Program Location	
North America	33 (89.19)
Europe	4 (10.81)
Number of Sessions	
1	5 (13.51)
2-10	23 (62.16)
11-20	4 (10.81)
>20	5 (13.51)
Number of Hours	
<2	6 (16.22)
2-10	22 (59.46)
11-20	4 (10.81)
>20	5 (13.51)
Delivery Method	
In-person	35 (94.59)
Online	2 (5.41)
Program Setting	
Classroom	27 (72.97)
Small Group	6 (16.22)
Large Group	1 (2.70)
Online	2 (5.41)
Other	1 (2.70)
Program Facilitator	
Teachers	9 (24.32)
Program facilitators	10 (27.03)
Other professional staff (e.g., school counsellor, social worker)	11 (29.73)
Online	2 (5.41)
Peer	3 (8.11)
Other	2 (5.41)

We also coded various components of the program curricula; however, as noted previously this information was much less consistently reported. As such, the following information regarding program curriculum components only represents a fraction of the full sample. Based on the available information, 11 programs specifically addressed the topic of gender roles and stereotypes (32%), 20 programs involved group activities (74%), 23 programs included discussions (82%), 17 incorporated role-play activities (65%), 11 included bystander training (37%), and 30 programs focused on skill development (86%).

Table 4-3. Program components of the included studies (n=37)

Characteristics	*n* (%)
Addressed Gender Roles	
No	23 (67.65)
Yes	11 (32.35)
Missing	3
Incorporated Bystander Training	
No	19 (63.33)
Yes	11 (36.67)
Missing	7
Included Focus on Skill Development	
No	5 (14.29)
Yes	30 (85.71)
Missing	2
Included Role Play Activities	
No	9 (34.62)
Yes	17 (65.38)
Missing	11
Included Group Activities	
No	7 (25.93)
Yes	20 (74.07)
Missing	10
Included Discussion	
No	5 (17.86)
Yes	23 (82.14)
Missing	9

Sample

Treatment group sample size at pretest ranged from 20 to 1,389 participants (*M* = 405.95, *SD* = 387.76), and treatment group sample size at post-test ranged from 20 to 1,372 (*M* = 374.10, *SD* = 344.27). Participants ranged in age from 10 to 25 years, with an approximate mean age of 14.52 (*SD* = 1.2). The samples were primarily mixed in gender, with 78% of studies including both males and females in the analytic sample, while 14% were female-only samples, and 8% targeted only males. Thirty-five percent of the studies had a predominantly Caucasian or mixed ethnicity sample, while 65% had a sample that was primarily composed of ethnic minorities. Additionally, six of the studies identified the sample as being at-risk for dating violence, with 26 samples identified as not at-risk (five studies had missing data).

Table 4-4. Sample characteristics of the included studies (n = 37)

Study Characteristics	*n* (%)
Treatment group sample size at pre-test	*M* = 405.95 (*SD* = 387.76)
Treatment group sample size at post-test	*M* = 374.10 (*SD* = 344.27)
Participant age	*M* = 14.52 (*SD* = 1.2)
Gender of Analysis Sample	
Mixed	29 (78.38)
Female only	5 (13.51)
Male only	3 (8.11)
Race/ethnicity of Analysis Sample	
Caucasian/Mixed	11 (35.48)
Minority	20 (64.52)
Missing	6
Sample "At-Risk"	
No	26 (81.25)
Yes	6 (18.75)
Missing	5

Research design

All 37 included studies employed a pretest-post-test design, however the exact study designs varied. Over half of the included studies used a randomized control trial design (57%) and 8% used a quasi-experimental design with a strongly matched comparison group. The remaining studies included quasi-experiments with a weakly matched comparison group (5%) and single group pretest-post-test designs (30%). Additionally, studies assigned participants to their respective treatment condition at the individual (24), classroom (30%), or school (41%) level[10]. Given the unit of assignment, many of the included studies had a clustered research design and as such, were cluster-adjusted for analysis (57%)[11]. There were also varying levels of researcher involvement; in 70% of the studies the researcher was only involved with the evaluation of the program, in 16% the researcher was also involved with the planning and supervision of

[10] Information regarding unit of assignment was missing for four studies. One study (Jaycox et al., 2006) assigned conditions based on "tracks" within a school, which equated to subgroups within a school that operated on different schedules.

[11] Studies cluster-adjusted at the classroom level were tested using both an ICC of .10 and .05, as well as with fixed- and random effects models. There were no substantive impacts on conclusions between models; see the discussion on cluster adjustments in the Method, section 3.4.3.

program implementation, in 5% the researcher directly delivered the program, and in the final 8% of studies the researcher was also the program developer.

Table 4-5. Research design characteristics of the included studies (n = 37)

Study Characteristics	*n* (%)
Research Design	
Randomized control trial	21 (56.76)
Quasi-experiment with matched comparison group	3 (8.11)
Quasi-experiment with weakly matched comparison group	2 (5.41)
Single group pre-test-post-test	11 (29.73)
Unit of Assignment	
"Tracks" Within a school	1 (2.70)
Classroom	11 (29.73)
Individual	9 (24.32)
School	15 (40.54)
Missing	1 (2.70)
Cluster Adjusted	
No	16 (43.24)
Yes	21 (56.76)
Level of Researcher Involvement	
Research Only	26 (70.27)
Planning /Supervising Intervention	6 (16.22)
Delivering Program Directly	2 (5.41)
Developed Program	3 (8.11)

Outcomes

The included studies used varying post-test and follow-up time points, with potential measures ranging from immediately after the program ended to four years after the program end. As discussed in section 3.4.5, when multiple follow-up periods were available, the immediate post-test or post-test closest to the end of the program was chosen, except when necessary statistical data were only presented for a later follow-up measurement. This resulted in the use of an immediate post-test for 51% of the included studies, 1-3 month follow-up for 27% of the studies, 4-6 month follow-up for 11% of the studies, and 7-month or longer follow-up for 11% of the studies.

Table 4-6. Outcome characteristics of the included studies (n = 37)

Study Characteristics	*n* (%)
Time of Post-test	
Immediate Post-test	19 (51.35)
1 – 3 months post-treatment end	10 (27.03)
4 – 6 months post-treatment end	4 (10.81)
7 – 12 months post-treatment end	1 (2.70)
12 months+	3 (8.11)

The following sections detail the results of the five outcomes measured. The outcomes include: 1) knowledge of dating violence, 2) attitudes and beliefs towards dating violence, 3) perpetration of dating violence behaviours, 4) experiences of dating violence victimization, and 5) bystander behaviours and intentions.

For each of the five outcomes, results pertaining to 1) main effects, 2) sensitivity analyses, and 3) moderator analyses are presented.

4.2. Outcome 1: Knowledge of dating violence

4.2.1. Pooled analysis

Outcomes examining the impact of the dating violence prevention programs on knowledge of dating violence were pooled together (n=16). Both fixed effects and random effects models were conducted, however, due to the significant (and expected) heterogeneity, the random effects model was selected[12]. A random effects model (Table 4.7) yielded a pooled estimate of 0.566 (z = 3.59, p < .001). This is a statistically significant, positive result which suggests that dating violence prevention programs are effective at increasing adolescents' knowledge of dating violence. In other words, youth are likely to demonstrate an increase in their knowledge about dating violence if they participate in a dating violence prevention program. The model resulted in a significant Q-statistic of 453.23 (df = 15, p < .001), suggesting there is significant heterogeneity within the sample; this heterogeneity can largely be attributed to factors beyond sampling error, as implied by the high I^2 value of 96.7%.

Table 4-7. Random effects meta-analysis for knowledge of dating violence

Pooled estimate	95% CI Lower	95% CI Upper	*Z* (*p*-value)	No. of Studies	*Q*-Statistic	I^2
0.566	0.257	0.875	3.59 (p<.001)	16	453.23, df=15, p<.001	96.7%

[12] Fixed effects and random effects models were conducted for all 5 outcomes, with random effects models selected and presented in the analyses.

Table 4-8. Study level data for the meta-analysis of the knowledge outcome

Author, year	Effect Size	95% CI Lower	95% CI Upper	Relative weight (%)
Silverman (1998)	0.504	0.376	0.633	6.51
Jaycox (2006)	0.706	0.563	0.849	6.50
Antle (2011)	0.914	0.744	1.084	6.46
Schramm (2012)	2.018	1.842	2.194	6.45
Miller (2012)	0.059	-0.141	0.259	6.41
Adler-Baeder (2007)	1.696	1.476	1.917	6.37
McLeod (2015)	1.125	0.902	1.349	6.37
Lowe (2015)	0.402	0.175	0.629	6.36
Miller (2015)	0.135	-0.131	0.401	6.28
Chamberland (2014)	0.395	0.119	0.671	6.26
Gardner (2004)	0.564	0.286	0.842	6.25
Joppa (2016)	-0.377	-0.736	-0.018	6.05
MacGowan (1997)	0.030	-0.337	0.397	6.03
Connolly (2015)	0.257	-0.115	0.629	6.01
DeGannes (2009)	0.403	-0.027	0.832	5.84
Sanchez-Cesareo (2002)	0.046	-0.385	0.477	5.84

Figure 4.1 below is a forest plot representing the random effects meta-analysis of the dating violence knowledge outcome. The forest plot visually represents the overall pooled effect, as well as the individual effect sizes for each study included in the analysis. The black diamond at the bottom of the plot represents the pooled effect, with the left and right corners representing the 95% confidence intervals. The individual black diamonds with the grey boxes correspond to the individual effect sizes, while the horizontal lines extending out from the box represent the 95% confidence intervals. The solid vertical line represents the “line of no effect” (i.e., the zero point), whereby if either the horizontal lines or the edges of the pooled effect diamond crosses this line, that effect is not statistically significant.

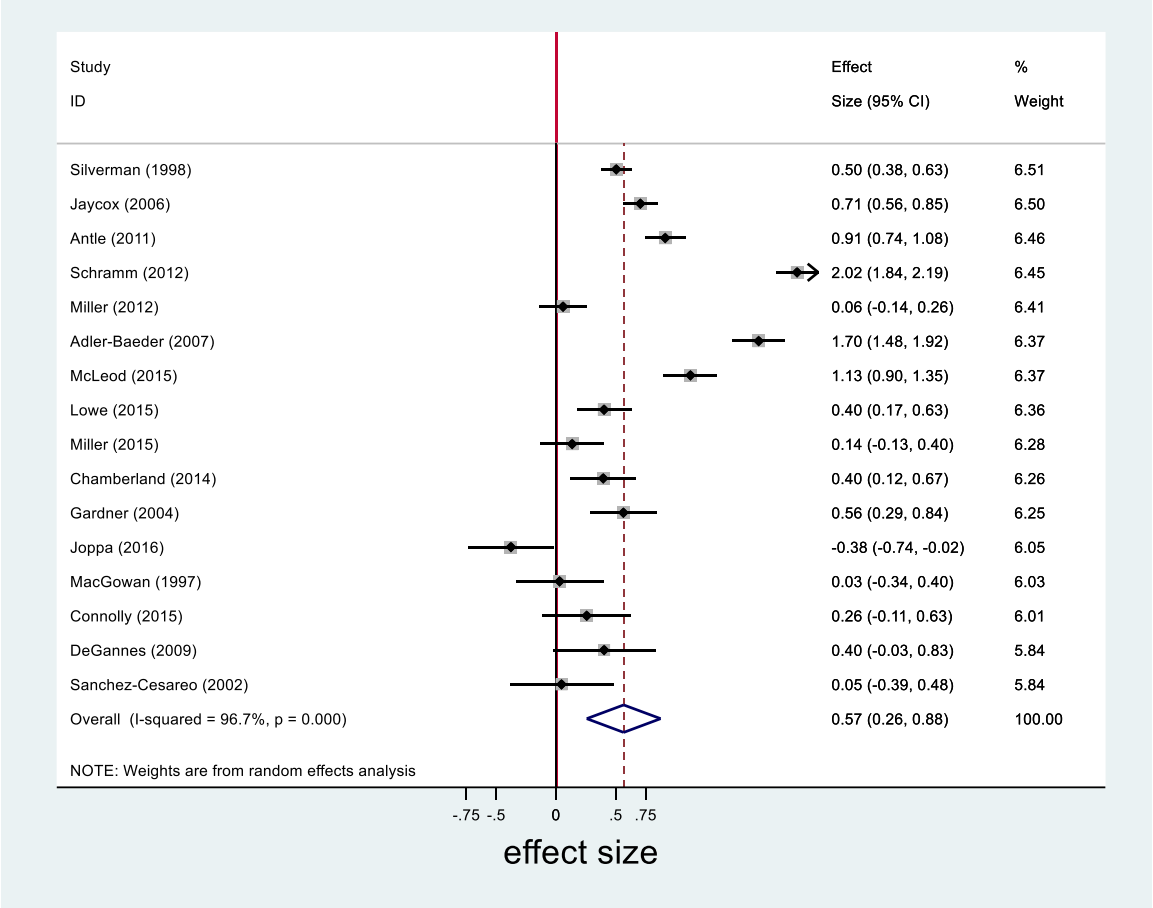

Figure 4-1. Forest plot for dating violence knowledge outcome

Examining Figure 4.1 the significant and positive pooled effect is visible at the bottom of the plot, to the right of the zero point. The plot also illustrates the significant level of heterogeneity between studies. However, we can see that only one effect size was to the left of the zero point, demonstrating that only one study produced a negative effect (Joppa, 2016[13]), though it was statistically significant (the confidence interval does not cross the zero point). Conversely, the remaining 15 studies appear to produce positive and mostly significant effects, with only six of these studies having effects that are not significant.

[13] Throughout the results presentation, the included studies will be referred to by the first author and date only (in tables, figures, and text) due to space constraints.

4.2.2. Sensitivity analysis

Publication bias

Bias was assessed using two methods: the funnel plot and Egger's test of small study effects. Figure 4.2 below depicts the funnel plot for dating violence knowledge and includes the Egger's regression line estimating the slope of the underlying effect; the plot depicts the individual study effect sizes and their relative standard errors within pseudo 95% confidence intervals. The estimates in the funnel plot show a considerable amount of asymmetry, with a gap in the lower right-hand section possibly indicating the presence of publication bias.

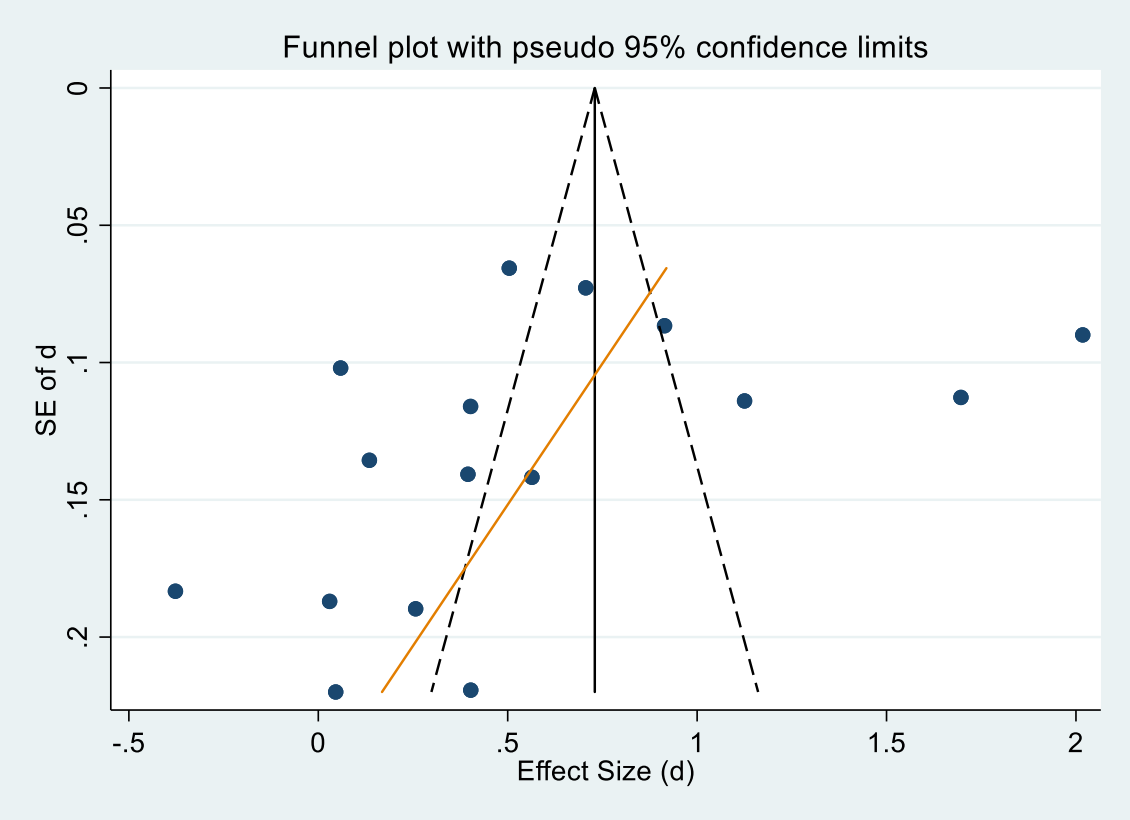

Figure 4-2. Funnel plot for dating violence knowledge

While the asymmetry in the funnel plot suggests the possibility of bias, Egger's test of small study effects provides no indication of bias with a coefficient of -4.863 (*SE* = 3.860, *t* = -1.26, *p* = .228). This result indicates the asymmetry illustrated in the funnel plot is not due to publication bias alone. Other potential sources of the between-study heterogeneity are explored through subgroup analysis.

Influence analysis

The funnel plot above also depicts multiple possible outliers which may be having a disproportionate influence on the pooled estimate. To test the individual influence of each separate study, influence analysis was conducted using random effects models[14]. Table 4.9 illustrates the pooled estimates if each individual study were removed from the analysis.

Table 4-9. Influence analysis on dating violence knowledge

Study Omitted	Effect Size	95% CI Lower	95% CI Upper
Adler-Baeder (2007)	0.490	0.189	0.792
Antle (2011)	0.541	0.204	0.878
Chamberland (2014)	0.577	0.252	0.902
Connolly (2015)	0.586	0.265	0.907
DeGannes (2009)	0.576	0.255	0.897
Gardner (2004)	0.566	0.240	0.892
Jaycox (2006)	0.555	0.209	0.901
Joppa (2016)	0.627	0.317	0.938
Lowe (2015)	0.577	0.249	0.904
MacGowan (1997)	0.601	0.282	0.919
McLeod (2015)	0.528	0.202	0.854
Miller (2012)	0.601	0.285	0.917
Miller (2015)	0.595	0.275	0.915
Sanchez-Cesareo (2002)	0.598	0.280	0.917
Schramm (2012)	0.473	0.232	0.713
Silverman (1998)	0.569	0.223	0.915
Combined	**0.566**	**0.257**	**0.875**

Examining the results, none of the included studies appear to have enough influence to change the outcome of the pooled estimate. While the overall pooled estimate does change considerably with the removal of some studies (e.g., Adler-Baeder, 2007; McLeod, 2015, or Schramm, 2012), none of them shift the pooled effect to a non-significant level. This suggests the overall effect of dating violence prevention programs on dating violence knowledge is robust. Figure 4.3 provides a visual representation of the results.

[14] For each of the 5 outcome categories, analyses were also conducted using fixed effects models. As no studies were identified that shifted the pooled estimate to a non-significant value, the random effects models are presented here to be consistent with the main effects models.

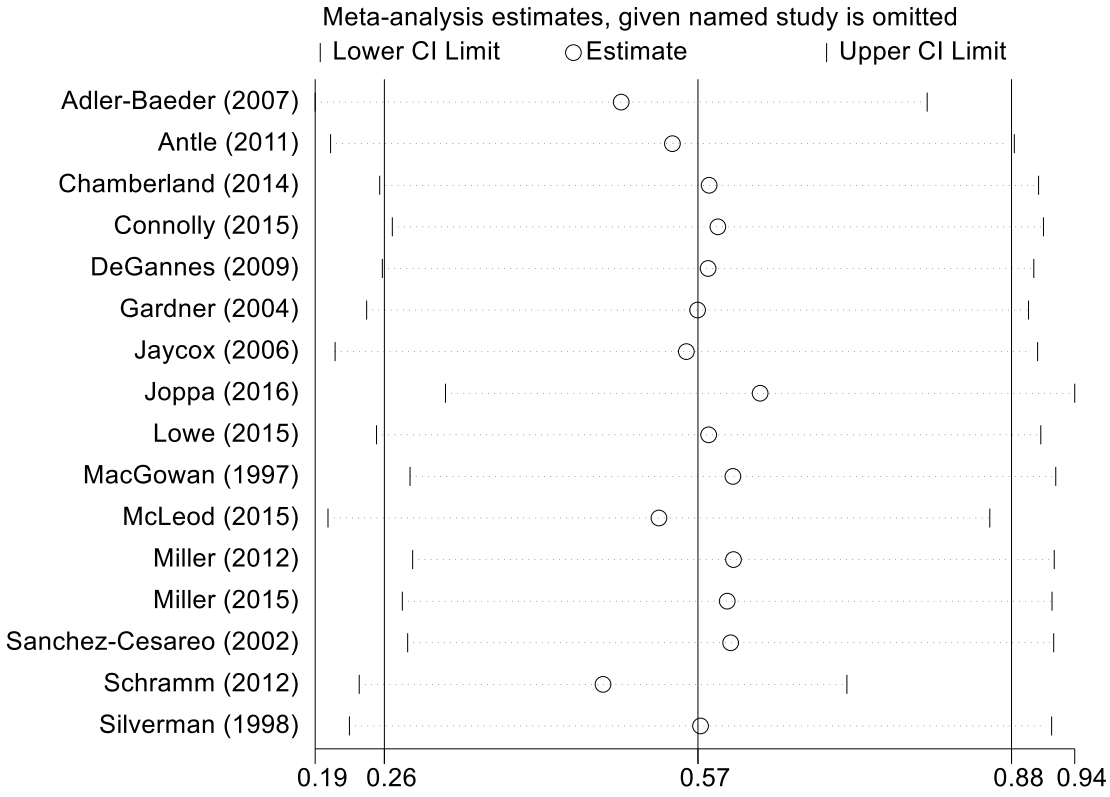

Figure 4-3. Influence analysis on dating violence knowledge

4.2.3. Assessing heterogeneity

The random effects model for the dating violence knowledge outcome resulted in a significant *Q*-statistic indicating a considerable amount of between-study heterogeneity. Given that the included studies evaluate different programs, a certain level of heterogeneity is expected. However, the I^2 suggests that 96.7% of the heterogeneity is due to factors other than sampling error. To assess the potential sources of heterogeneity, the analog to the ANOVA method was used to compare subgroups.

Using this method, a series of seven dichotomous program and study characteristics were selected[15] to investigate the heterogeneity. Program characteristics

[15] The specific variables chosen for analysis vary among the five outcome measure categories. This is due to differing sample sizes and whether the outcome had sufficient observations and variability between observations to conduct meaningful comparisons.

include: 1) length of program (≤ 5 hours versus ≥ 6 hours); and 2) curriculum addressed gender roles and stereotypes (yes/no).

Study characteristics include: 1) treatment group sample size (< 300 versus ≥ 300); 2) research design (randomized control trial versus non-randomized); 3) publication year (prior to 2015 versus 2015 or later); 4) time of post-test (immediately after program end versus 1+ months after program end); and 5) sample ethnicity (predominantly Caucasian or mixed versus predominantly an ethnic minority).

Table 4-10. Fixed effects moderator analysis for the dating violence knowledge outcome

Study characteristic	Effect size and *Q*-statistics
Length of program (hours) (n=16)	
5 or fewer hours (n=11)	*ES* = .452, z = 13.4, p < .001*
6 or more (n=5)	*ES* = 1.33, z = 26.9, p < .001*
Between-studies heterogeneity	Q_B = 215.67 ~χ^2_1, p < .001*
Within-studies heterogeneity	Q_W = 237.56 ~χ^2_{14}, p < .001*
Program addressed gender roles (n=14)[a]	
No (n=10)	*ES* = .946, z = 25.44, p < .001*
Yes (n=4)	*ES* = .380, z = 6.95, p < .001*
Between-studies heterogeneity	Q_B = 122.22 ~χ^2_1, p < .001*
Within-studies heterogeneity	Q_W = 331.01 ~χ^2_{12}, p < .001*
Treatment group sample size (n=16)	
Less than 300 (n=10)	*ES* = .694, *z* = 19.04, *p* < .001*
300 or more (n=6)	*ES* = .780, *z* = 18.03, *p* < .001*
Between-studies heterogeneity	Q_B = 2.31 ~χ^2_1, *p* = .129
Within-studies heterogeneity	Q_W = 450.92 ~χ^2_{14}, *p* < .001*
Randomized control trial (n=16)	
No (n=9)	*ES* = .957, *z* = 27.25, *p* < .001*
Yes (n=7)	*ES* = .342, *z* = 7.47, *p* < .001*
Between-studies heterogeneity	Q_B = 113.35 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 339.88 ~χ^2_{14}, *p* < .001*
Publication year (n=16)	
Prior to 2015 (n=12)	*ES* = .774, *z* = 25.63, *p* < .001*
2015 - 2019 (n=4)	*ES* = .476, *z* = 6.54, *p* < .001*
Between-studies heterogeneity	Q_B = 14.30 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 438.93 ~χ^2_{14}, *p* < .001*
Time of post-test (n=16)	
Immediate post-test (n=12)	*ES* = .790, *z* = 26.59, *p* < .001*
Post-test 1 month+ (n=4)	*ES* = .285, *z* = 3.52, *p* < .001*
Between-studies heterogeneity	Q_B = 34.25 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 418.98 ~χ^2_{14}, *p* < .001*
Sample ethnicity (n=15)[b]	
Caucasian/mixed (n=5)	*ES* = .963, *z* = 22.35, *p* < .001*
Predominant minority (n=10)	*ES* = .574, *z* = 15.15, *p* < .001*
Between-studies heterogeneity	Q_B = 51.88 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 401.35 ~χ^2_{13}, *p* < .001*

a. Two studies were missing from this analysis as information regarding whether the topic of gender roles was covered in the program curriculum was not provided.
b. One study was missing from this analysis as it did not provide information regarding the sample ethnicity.

As evidenced in Table 4.10 all tested characteristics, with the exception of treatment group sample size, were significant moderator variables. Although we were only able to examine two program characteristics, both were significant. The length of the program (in hours) was a significant moderator with a significant *Q*-between result of 215.67 ($p < .001$). The larger effect size for longer programs suggests that programs

longer in length are more effective at increasing adolescents' knowledge of dating violence. Additionally, whether a program's curriculum addressed gender roles was also a significant moderating variable. The *Q*-between statistic was significant (Q_B 122.22, $p <$ .001); as stated in Table 4.10, programs that do not address gender roles produced significantly larger effect sizes than those that do address gender roles, implying that programs covering this topic are less effective at increasing dating violence knowledge.

All but one of the study characteristic variables were also significant in the analyses and, thus, are significant explanatory factors of the between study heterogeneity. Research design was an important moderating variable (Q_B = 113.35, $p <$ .001), with studies using a randomized design producing significantly smaller effect sizes than those that used a non-randomized design (e.g., single group pretest-post-tests and quasi-experiments with match comparison groups). This finding is consistent with literature that has found non-randomized or 'weaker' designs, such as single group pretest-post-test designs, tend to overestimate the treatment effect (Lipsey & Wilson, 2001; Morris & DeShon, 2002).

Publication year also resulted in a significant *Q*-between statistic (14.30, $p <$ .001), with studies that were published prior to the year 2015 producing significantly larger effect sizes than studies published in 2015 or later. Additionally, the time of post-test was a significant moderator. This analysis resulted in a significant *Q*-between value of 34.25 ($p <$ .001), with measures at an immediate post-test producing significantly larger effect sizes than those with measures at least one-month after program end. This result fits with expectations, as program effects are expected to be strongest at the immediate conclusion of a program and then dissipate over time.

Sample characteristics were also examined and results show that sample ethnicity is a significant moderator and helps explain between-study heterogeneity. The *Q*-between statistic was significant at 51.88 ($p <$ .001); the significant effect sizes for each group suggest that studies with predominantly Caucasian or mixed ethnicity samples produce significantly larger effects sizes than those with samples composed primarily of ethnic minorities. Given that a majority of programs were likely developed for a general population, this finding is consistent with previous research which has found that adult domestic violence programs are more effective with minority populations when they are culturally-focused (Babcock et al., 2016).

Treatment group sample size was the only non-significant moderator, with studies with samples under 300 participants producing equally significant effect sizes as studies with samples larger than 300 participants. However, it is important to note that a sample size of 300 participants is quite large; results may have differed if the comparison sample size was a smaller value.

4.3. Outcome 2: Attitudes and beliefs towards dating violence

4.3.1. Pooled analysis

The included studies produced a set of 20 effect sizes measuring the outcome of attitudes and beliefs towards dating violence, which were pooled together for an overall effect estimate. A random effects model resulted in a statistically significant, positive effect of 0.191 (z = 3.88, p < .001) – see Table 4.11. These results suggest that dating violence prevention programs are effective at improving adolescents' attitudes and beliefs regarding dating violence and that youth who participate in these programs are less likely to be accepting of dating violence behaviours. The statistically significant Q-statistic (60.67, df = 19, p < .001) illustrates a considerable amount of heterogeneity within the sample, of which a majority (I^2 = 68.7%) can be attributed to factors other than sampling error.

Table 4-11. Random effects meta-analysis for attitudes towards dating violence

Pooled estimate	95% CI Lower	95% CI Upper	Z (*p*-value)	No. of Studies	*Q*-Statistic	I^2
0.191	0.094	0.287	3.88 (*p*<.001)	20	60.67, df=19, *p*<.001	68.7%

Table 4-12. Study level data for the meta-analysis of the attitude towards dating violence outcome

Study	Effect Size	95% CI Lower	95% CI Upper	Relative weight (%)
Silverman (1998)	0. 313	0.192	0.433	7.63
Jaycox (2006)	-0.030	-0.171	0.111	7.30
Levesque (2016)	0.141	-0.003	0.286	7.25
Antle (2011)	0.125	-0.024	0.274	7.18
Belknap (2013)	0.107	-0.051	0.265	7.03
Lowe (2015)	0.062	-0.115	0.238	6.72
Miller (2012)	0.000	-0.185	0.185	6.57
Southgate (2016)	-0.049	-0.269	0.170	5.98
Chamberland (2014)	0.332	0.057	0.607	5.08
Adler-Baeder (2007)	0.079	-0.202	0.360	5.00
de Graaf (2016)	0.467	0.178	0.756	4.88
McLeod (2015)	0.946	0.635	1.258	4.55
Joppa (2016)	0.539	0.178	0.900	3.91
Connolly (2015)	-0.039	-0.410	0.332	3.79
Rizzo (2019)	0.000	-0.377	0.377	3.73
Rizzo (2018)	0.225	-0.153	0.603	3.72
DeGannes (2009)	0.417	0.009	0.825	3.40
Banyard (2019)	0.352	-0.212	0.916	2.19
Avery-Leaf (1997)	0.255	-0.313	0.824	2.16
Miller (1998)	-0.041	-0.653	0.572	1.93

As illustrated in the forest plot below (Figure 4.4), the pooled effect is on the right-hand side of the plot and does not cross the zero point, visually confirming a positive and significant effect. The high level of heterogeneity is also visible among the individual effect sizes. Of the 20 individual studies, four are negative (though statistically insignificant), two had no effect, and 14 are positive (with eight having non-significant results).

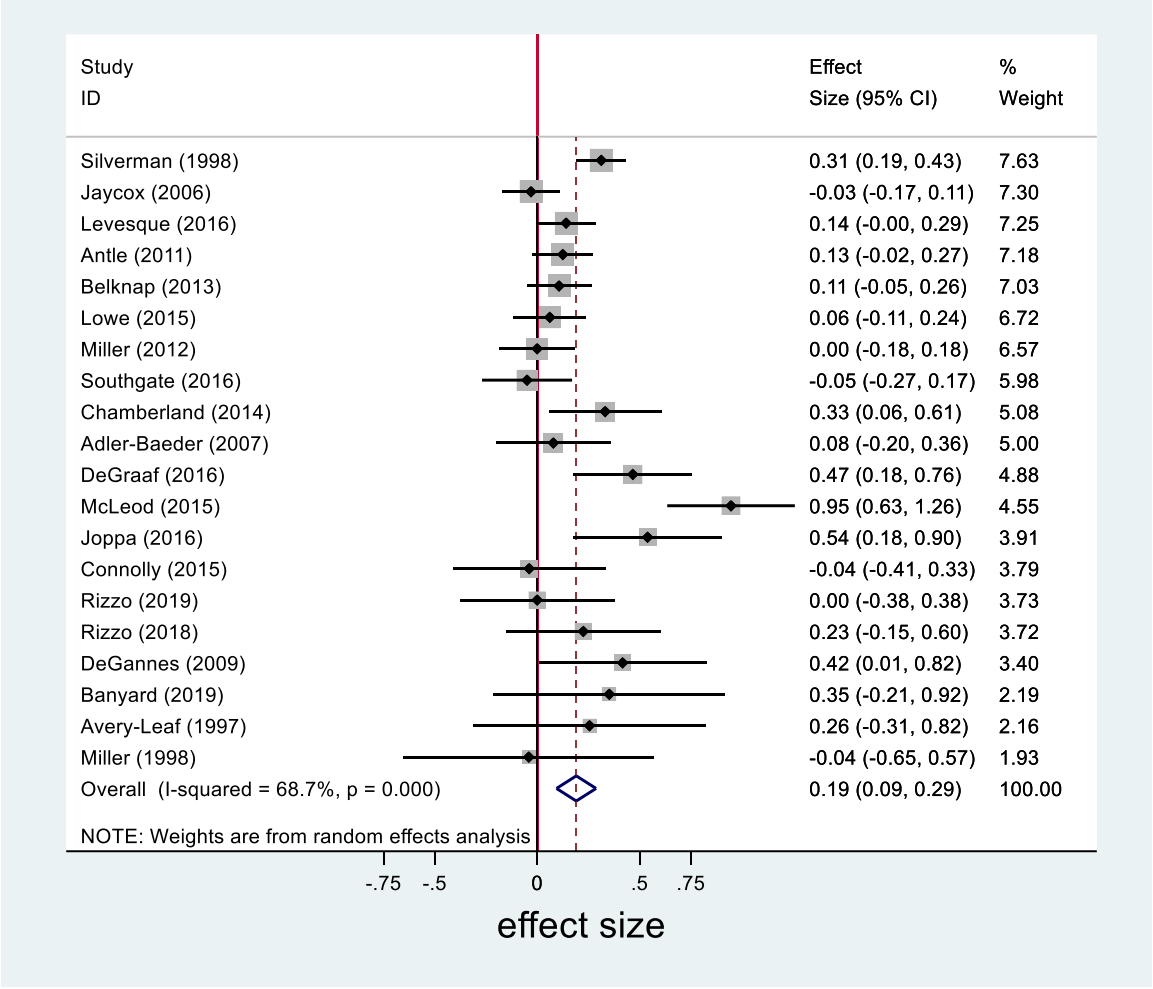

Figure 4-4. Forest plot for attitudes towards dating violence outcome

4.3.2. Sensitivity analysis

Publication bias

The funnel plot presented in Figure 4.5 appears to be relatively symmetrical, with only a few potential outliers. Specifically, there is one effect size in particular that fell quite far outside the pseudo 95% confidence intervals; this estimate will be investigated in detail using influence analysis. Other than the one extreme value, the majority of the estimates appear to fall mostly evenly throughout the funnel, providing support for a lack of publication bias.

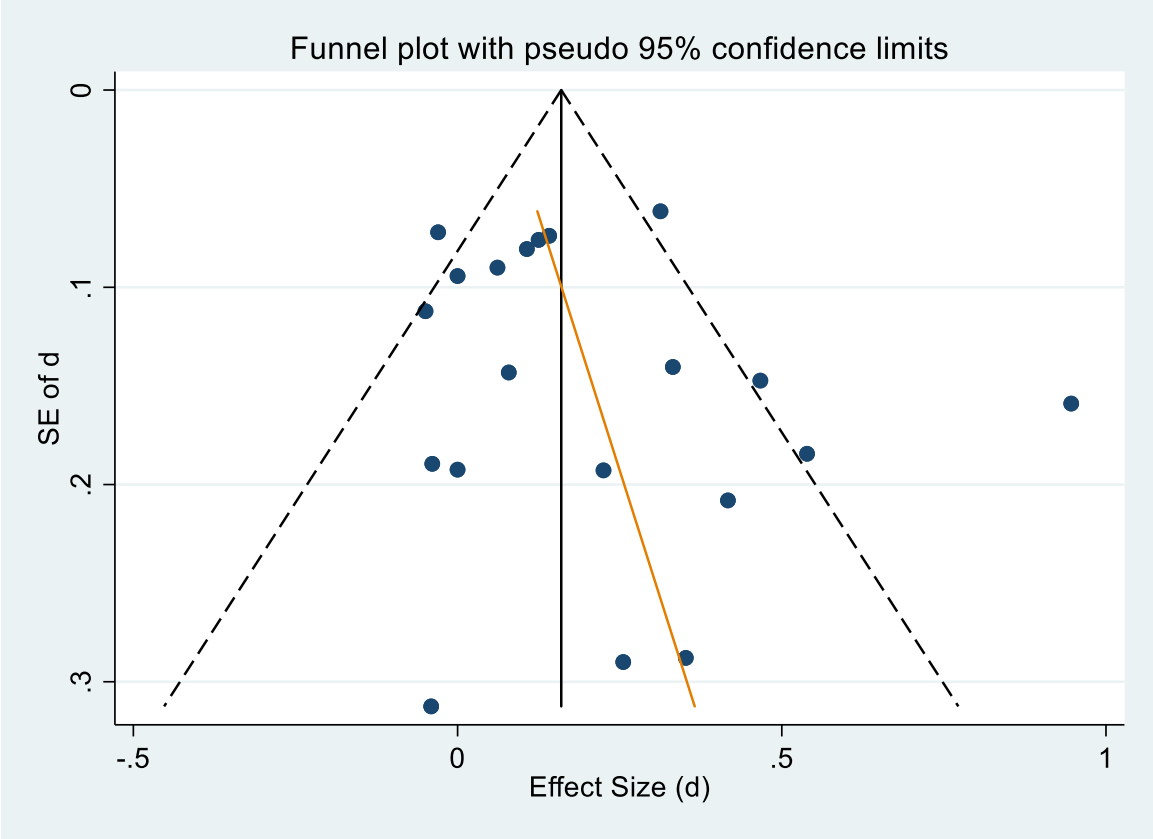

Figure 4-5. Funnel plot for attitudes towards dating violence

Egger's test of small study effects was conducted for attitudes towards dating violence and the results are consistent with the funnel plot. Egger's bias coefficient of 0.969 (*SE* = .920, *t* = 1.05, *p* = .307) was not statistically significant, indicating small study effects are not a concern.

Influence analysis

As depicted in the funnel plot, a few potential outliers fall outside the pseudo 95% confidence intervals, of particular note is the estimate on the far right side of the graph (McLeod, 2015). Table 4.13 and Figure 4.6 show the results of the influence analysis. The results indicate that the removal of McLeod (2015) affects the pooled estimate considerably, however, the removal does not shift the pooled estimate to a non-statistically significant outcome. This suggests the pooled effect is robust; no studies were removed due to outlier effects.

Table 4-13. Influence analysis on attitudes towards dating violence

Study Omitted	Effect Size	95% CI Lower	95% CI Upper
Adler-Baeder (2007)	0.198	0.097	0.298
Antle (2011)	0.198	0.093	0.303
Avery-Leaf (1997)	0.190	0.091	0.288
Banyard (2019)	0.188	0.089	0.286
Belknap (2013)	0.199	0.095	0.303
Chamberland (2014)	0.184	0.084	0.284
Connolly (2015)	0.200	0.101	0.299
DeGannes (2009)	0.183	0.085	0.281
de Graaf (2016)	0.176	0.079	0.273
Jaycox (2006)	0.208	0.109	0.307
Joppa (2016)	0.176	0.080	0.272
Levesque (2016)	0.197	0.092	0.302
Lowe (2015)	0.201	0.099	0.304
McLeod (2015)	0.147	0.070	0.224
Miller (1998)	0.196	0.098	0.294
Miller (2012)	0.205	0.104	0.306
Rizzo (2019)	0.199	0.099	0.298
Rizzo (2018)	0.190	0.090	0.290
Silverman (1998)	0.181	0.080	0.282
Southgate (2016)	0.206	0.107	0.305
Combined	**0.191**	**0.094**	**0.288**

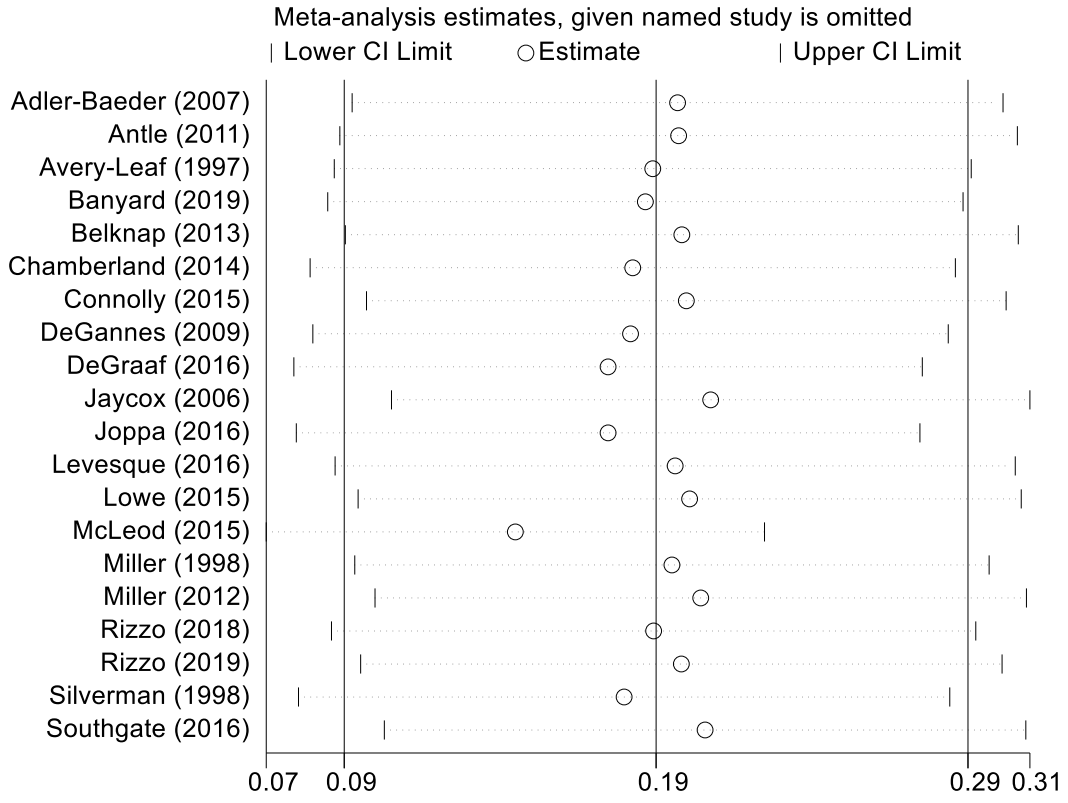

Figure 4-6. Influence analysis for attitudes towards dating violence

4.3.3. Assessing heterogeneity

Moderator analysis was also conducted to assess the heterogeneity of the attitudes towards dating violence outcome category. As the random effects model showed significant heterogeneity with a statistically significant *Q*-statistic (60.67, *df* = 19, $p < .001$), these analyses aimed to determine if systematic heterogeneity was present. The following analyses examined the same seven program and study characteristics as for the knowledge outcome moderator analyses, with the addition of the school level variable, which measured whether the program was implemented in a middle school or high school. The characteristics include: length of program (≤ 5 hours versus ≥ 6 hours), whether the curriculum addressed gender roles (yes/no), what level of school the program was implemented in (middle school versus high school), treatment group sample size (<300 versus ≥300), research design (randomized control trial versus non-randomized), publication year (prior to 2015 or 2015 and later), time of post-test

(immediately after program end versus 1+ months after program end), and sample ethnicity (predominantly Caucasian or mixed versus predominantly an ethnic minority).

Table 4.14 demonstrates the results of all eight moderator variables across all studies within the attitudes towards dating violence outcome category (*n* = 20). Only one of the program variables and three of the study variables were significant moderators, suggesting that a majority of these variables are not important at moderating program effects or explaining the heterogeneity between studies.

Table 4-14. Fixed effects moderator analysis for the attitudes towards dating violence outcome

Study characteristic	Effect size and *Q*-statistics
Length of program (hours) (n=20)	
5 or fewer hours (n=13)	*ES* = .166, *z* = 5.80, *p* < .001*
6 or more (n=7)	*ES* = .143, *z* = 2.89, *p* < .01*
Between-studies heterogeneity	Q_B = 0.15 ~χ^2_1, *p* = .699
Within-studies heterogeneity	Q_W = 60.52 ~χ^2_{18}, *p* < .001*
Program addressed gender roles (n=17)[a]	
No (n=11)	*ES* = .104, *z* = 3.16, *p* < .01*
Yes (n=6)	*ES* = .323, *z* = 6.29, *p* < .001*
Between-studies heterogeneity	Q_B = 20.41 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 40.26 ~χ^2_{15}, *p* < .05*
Level of school (n=19)[b]	
Middle school (n=6)	*ES* = .113, *z* = 1.81, *p* = .071
High school (n=13)	*ES* = .175, *z* = 6.07, *p* < .001*
Between-studies heterogeneity	Q_B = 1.02 ~χ^2_1, *p* = .313
Within-studies heterogeneity	Q_W = 59.65 ~χ^2_{17}, *p* < .0001*
Treatment group sample size (n=20)	
Less than 300 (n=16)	*ES* = .205, *z* = 6.78, *p* < .001*
300 or more (n=4)	*ES* = .069, *z* = 1.59, *p* = .111
Between-studies heterogeneity	Q_B = 6.69 ~χ^2_1, *p* < .05*
Within-studies heterogeneity	Q_W = 53.98 ~χ^2_{18}, *p* < .001*
Randomized control trial (n=20)	
No (n=10)	*ES* = .209, *z* = 6.48, *p* < .001*
Yes (n=10)	*ES* = .090, *z* = 2.33, *p* < .05*
Between-studies heterogeneity	Q_B = 5.59 ~χ^2_1, *p* < .05*
Within-studies heterogeneity	Q_W = 55.08 ~χ^2_{18}, *p* < .001*
Publication year (n=20)	
Prior to 2015 (n=11)	*ES* = .134, *z* = 4.56, *p* < .001*
2015 - 2019 (n=9)	*ES* = .225, *z* = 4.88, *p* < .001*
Between-studies heterogeneity	Q_B = 2.79 ~χ^2_1, *p* = .095
Within-studies heterogeneity	Q_W = 57.88 ~χ^2_{18}, *p* < .001*
Time of post-test (n=20)	
Immediate post-test (n=14)	*ES* = .156, *z* = 5.66, *p* < .001*
Post-test 1 month+ (n=6)	*ES* = .175, *z* = 3.15, *p* <.01*
Between-studies heterogeneity	Q_B = 0.09 ~χ^2_1, *p* = .764
Within-studies heterogeneity	Q_W = 60.58 ~χ^2_{18}, *p* < .001*
Sample ethnicity (n=18)[c]	
Caucasian/mixed (n=9)	*ES* = .192, *z* = 5.49, *p* < .001*
Predominant minority (n=9)	*ES* = .111, *z* = 3.05, *p* < .01*
Between-studies heterogeneity	Q_B = 4.66 ~χ^2_1, *p* < .05*
Within-studies heterogeneity	Q_W = 56.01 ~χ^2_{16}, *p* < .001*

a. Three studies are missing from this model as they did not specify whether or not gender roles was a topic included in the program curriculum.
b. One study is missing from this model as it did not specify whether the program was implemented in middle or high schools.
c. Two studies are missing from this model as information on sample ethnicity was not provided.

When examining the program characteristics for which sufficient information was available to conduct moderator analyses, neither program length (hours) or level of school were found to be significant moderating variables. No differences in treatment effects were found between shorter and longer programs (Q_B = 0.15, p = .699). However, both shorter (ES = .166, z = 5.80, p < .001) and longer (ES = .143, z = 2.89, p < .01) programs had significant effect sizes. This suggests that any dating violence prevention program, regardless of program length, is effective at improving attitudes towards dating violence. As well, the level of school in which the program was implemented was also not a significant moderator, with no differences between the effect sizes produced by programs implemented in middle schools or high schools.

Whether a program's curriculum addressed gender roles was the only significant moderating variable within the program characteristics (Q_B = 20.41, p < .001). The effect sizes suggest programs that include information regarding gender roles in the curriculum produce significantly larger effect sizes than those that do not. It is also important to note that both categories of programs resulted in significant effect sizes, which demonstrates that while programs with gender role-specific content produce larger effect sizes, programs that do not specific address gender roles also have a positive impact on attitudes towards dating violence. This finding suggests that incorporating gender role related content is not required for improving adolescent attitudes, but may enhance the effect if included.

Several of the study characteristic variables were also significant moderating variables, one of which was treatment group sample size. The Q-between statistic was 6.69 (p < .05), meaning studies with smaller sample sizes (< 300 participants) produced significantly larger effect sizes. This finding is consistent with literature that suggests small sample size can lead to inaccurate estimates of treatment effect due to increased sampling error (Lipsey & Wilson, 2000). Additionally, research design was significant (Q_B = 5.59, p < .05), with effect sizes demonstrating that non-randomized research designs produce significantly larger effect sizes than randomized designs. This is similar to the results of the moderator analysis for the dating violence knowledge outcome, and is also consistent with previous literature which indicates that weaker methodological designs tend to overestimate the treatment effect (Carlson & Schmidt, 1999). As such, we would expect more methodologically rigorous studies to elicit smaller effect sizes. Publication

year was not a statistically significant moderator, with a non-significant *Q*-between statistic of 2.79 ($p = 0.095$).

Again, similar to the dating violence knowledge outcome, sample ethnicity was a significant moderating variable with a *Q*-between statistic of 4.66 ($p < .05$), suggesting that studies with samples primarily comprised of Caucasian participants or participants from a variety of ethnic backgrounds produced significantly larger effect sizes than those with samples comprised predominantly of ethnic minorities.

The timing of the post-test was not found to be a significant moderator, with no statistically significant differences between those studies with an immediate post-test and those with a measure one month or more past the end of the program ($Q_B = 0.09$, $p = .764$).

4.4. Outcome 3: Perpetration of dating violence behaviours

4.4.1. Pooled analysis

Effect sizes (n=16) estimating the effect of dating violence prevention programs on the perpetration of violent behaviours were pooled together using a random effects model. The pooled effect was significant at 0.156 ($z = 2.94$, $p < .01$), suggesting a positive impact of prevention programs on reducing incidents of dating violence behaviours. These results demonstrate that incidents of dating violence among adolescents can be reduced and prevented with the implementation of dating violence prevention programs. The model did result in a significant *Q*-statistic of 42.28 ($df = 15$, $p < .001$), which indicates a significant amount of heterogeneity among the sample. The I^2 value of 64.5% suggests that a majority of the heterogeneity can be attributed to factors outside of sampling error.

Table 4-15. Random effects meta-analysis for incidents of dating violence perpetration

Pooled estimate	95% CI Lower	95% CI Upper	Z (*p*-value)	No. of Studies	*Q*-Statistic	I^2
0.156	0.052	0.260	2.94 (*p*<.01)	16	42.28, df=15, *p*<.001	64.5%

Table 4-16. Study level data for the meta-analysis of the incidence of dating violence perpetration outcome

Author, year	Effect Size	95% CI Lower	95% CI Upper	Relative weight (%)
Reidy (2017)	0.025	-0.105	0.156	9.67
Sanchez-Jiminez (2018)	-0.052	-0.188	0.084	9.55
Levesque (2016)	0.475	0.332	0.617	9.41
Schramm (2012)	0.251	0.058	0.444	8.20
Miller (2012)	0.102	-0.103	0.307	7.92
Ball (2012)	0.076	-0.157	0.309	7.25
Peskin (2014)	0.221	-0.017	0.458	7.16
Edwards (2019)	0.064	-0.191	0.319	6.77
Gardner (2004)	0.174	-0.096	0.444	6.46
Adler-Baeder (2007)	0.271	-0.014	0.556	6.15
Cissner (2014)	-0.124	-0.458	0.210	5.26
de Graaf (2016)	0.406	0.064	0.749	5.11
Muñoz-Fernández	0.032	-0.375	0.439	4.16
Rizzo (2019)	0.011	-0.512	0.534	2.94
Gonzalez-Guarda (2015)	0.511	-0.023	1.044	2.86
Wolfe (2009)	0.131	-0.795	1.057	1.14

The forest plot depicted in Figure 4.7 shows the significant and positive pooled effect. The plot also illustrates the heterogeneity within the sample; however it is evident that while the sample is heterogeneous, it is less so than for the previous two outcomes. Two of the individual studies produced negative (and non-significant) effects, while the remaining studies produced positive effects. However, only three of the positive effects are statistically significant, with six of the studies producing an effect size of 0.10 or smaller.

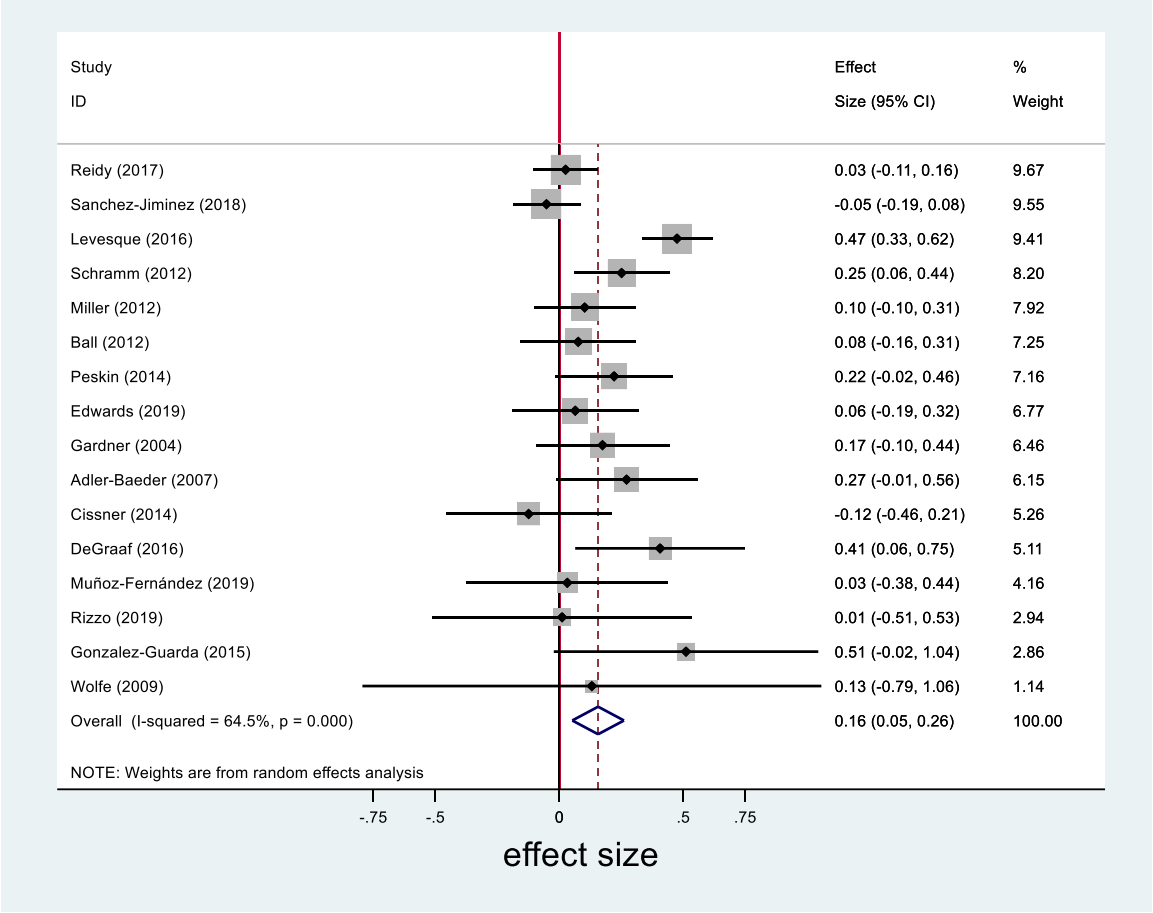

Figure 4-7. Forest plot for the violence perpetration outcome

4.4.2. Sensitivity analysis

Publication bias

To assess the potential presence of publication bias, a visual inspection of the funnel plot (Figure 4.8) was conducted. There is a slight asymmetry suggesting the possibility of publication bias. Egger's test of small study effects was conducted, and the bias coefficient was not statistically significant (0.281, *SE* = 1.017, *t* = 0.28, *p* = 0.787), indicating publication bias is unlikely.

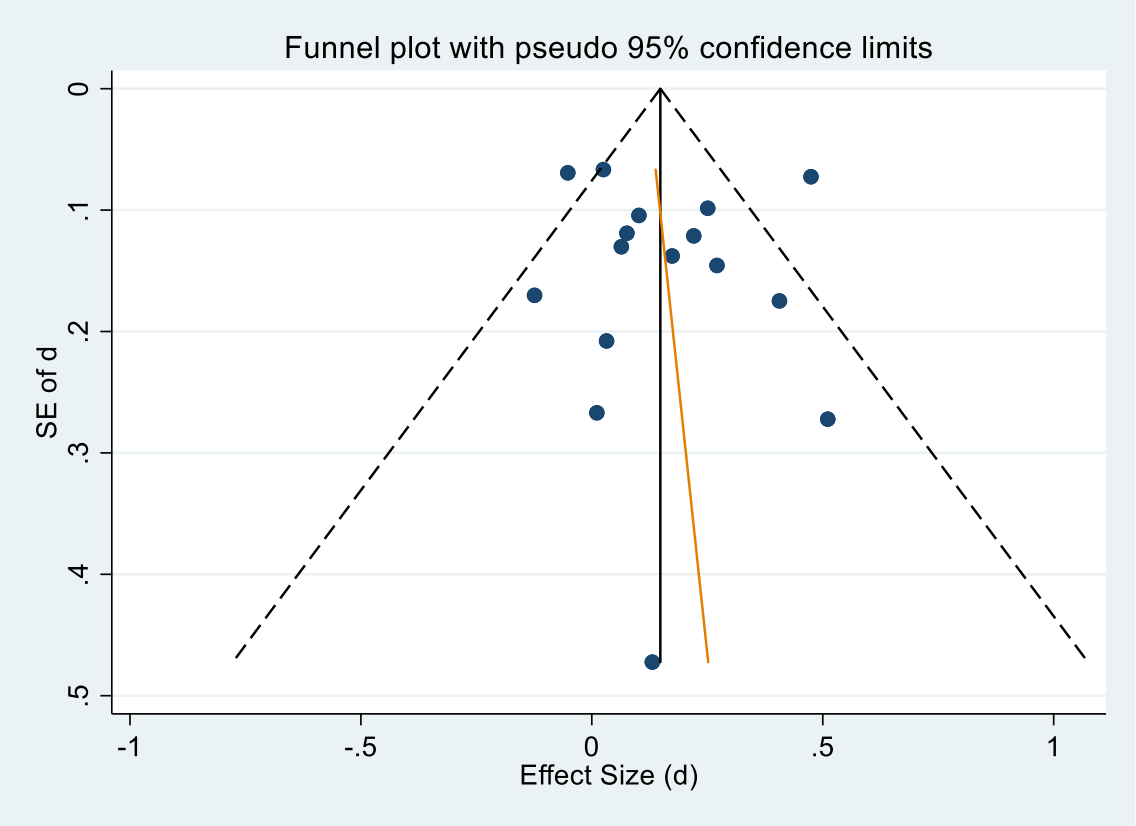

Figure 4-8. Funnel plot for incidents of violence perpetration

Influence analysis

Visual inspection of the funnel plot in Figure 4.8 identified one potential outlier (Levesque, 2016). Influence analysis was conducted to evaluate the influence of each individual study (Table 4.17 and Figure 4.9). Examining the results of the influence analysis, we see that the removal of Levesque (2016) does shift the overall pooled estimate down considerably, suggesting that the Levesque study is pulling the estimate up. However, this shift does not reach a non-statistically significant level, implying that retaining Levesque (2016) in the analysis is not problematic in terms of the overall analytic conclusion.

Table 4-17. Influence analysis for incidents of dating violence perpetration

Study Omitted	Effect Size	95% CI Lower	95% CI Upper
Adler-Baeder (2007)	0.149	0.039	0.258
Ball (2012)	0.163	0.051	0.274
Cissner (2014)	0.172	0.066	0.278
de Graaf (2016)	0.143	0.036	0.250
Edwards (2019)	0.163	0.052	0.274
Gardner (2004)	0.155	0.044	0.266
Gonzalez-Guarda (2015)	0.146	0.041	0.251
Levesque (2016)	0.106	0.031	0.181
Miller (2012)	0.161	0.048	0.274
Munoz-Fernandez (2019)	0.162	0.054	0.270
Peskin (2014)	0.151	0.040	0.263
Reidy (2017)	0.170	0.058	0.283
Rizzo (2019)	0.161	0.054	0.268
Sanchez-Jiminez (2018)	0.179	0.076	0.282
Schramm (2012)	0.148	0.036	0.260
Wolfe (2009)	0.157	0.050	0.263
Combined	**0.156**	**0.052**	**0.260**

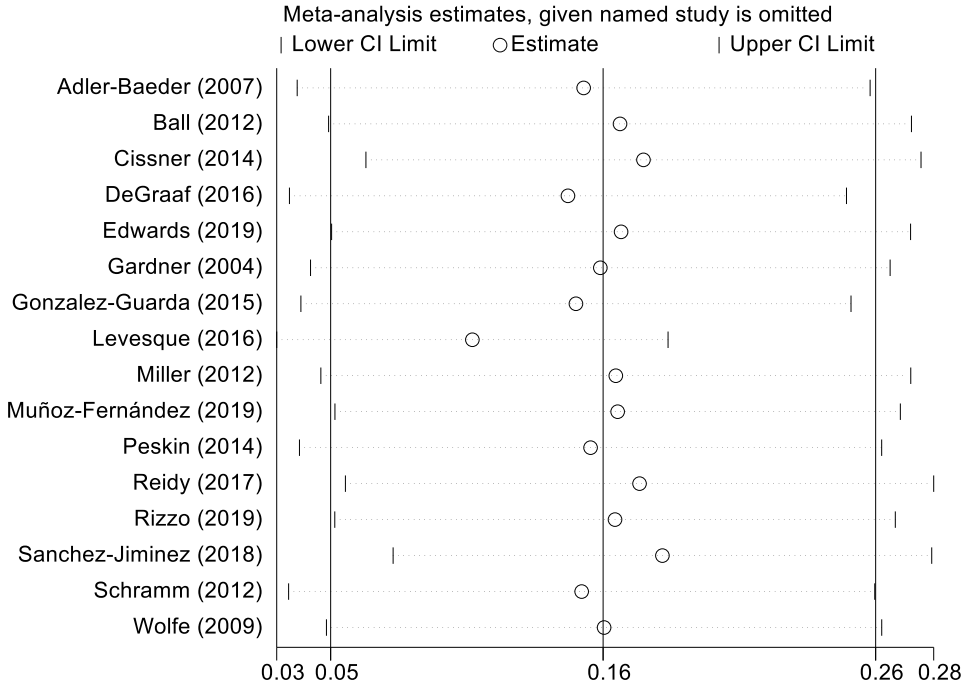

Figure 4-9. Influence analysis for incidents of violence perpetration

4.4.3. Assessing heterogeneity

To assess heterogeneity in studies for outcomes related to incidents of violence perpetration, two program characteristics and five study characteristics were included as potential moderator variables: whether the curriculum addressed gender roles (yes/no); whether the curriculum included bystander training (yes/no); treatment group sample size (< 300 versus ≥ 300); research design (randomized control trial versus non-randomized); publication year (prior to 2015 versus 2015 or later); time of post-test (immediately after program end versus 1+ months after program end); and sample ethnicity (predominantly Caucasian or mixed versus predominantly an ethnic minority).

Table 4-18. Fixed effects moderator analysis for incidents of dating violence perpetration

Study characteristic	Effect size and *Q*-statistics
Program addressed gender roles (n=14)[a]	
No (n=10)	*ES* = .192, *z* = 5.32, *p* < .001*
Yes (n=4)	*ES* = .040, *z* = 0.74, *p* = .462
Between-studies heterogeneity	Q_B = 7.88 ~χ^2_1, *p* < .01*
Within-studies heterogeneity	Q_W = 34.4 ~χ^2_{12}, *p* = .06
Bystander (n=14)[b]	
No (n=8)	*ES* = .231, *z* = 5.83, *p* < .001*
Yes (n=6)	*ES* = .013, *z* = 0.27, *p* = .785
Between-studies heterogeneity	Q_B = 13.70 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 28.58 ~χ^2_{12}, *p* = .195
Treatment group sample size (n=16)	
Less than 300 (n=6)	*ES* = .207, *z* = 3.14, *p* < .05*
300 or more (n=10)	*ES* = .135, *z* = 4.29, *p* < .001*
Between-studies heterogeneity	Q_B = 0.96 ~χ^2_1, *p* = .327
Within-studies heterogeneity	Q_W = 41.32 ~χ^2_{14}, *p* < .05*
Randomized control trial (n=16)	
No (n=6)	*ES* = .136, *z* = 3.13, *p* < .01*
Yes (n=10)	*ES* = .158, *z* = 4.20, *p* < .001*
Between-studies heterogeneity	Q_B = 0.16 ~χ^2_1, *p* = .689
Within-studies heterogeneity	Q_W = 42.12 ~χ^2_{14}, *p* < .05*
Publication year (n=16)	
Prior to 2015 (n=8)	*ES* = .156, *z* = 3.40, *p* < .01*
2015 - 2019 (n=8)	*ES* = .143, *z* = 3.97, *p* < .01*
Between-studies heterogeneity	Q_B = 0.05 ~χ^2_1, *p* = .823
Within-studies heterogeneity	Q_W = 42.23 ~χ^2_{14}, *p* < .05*
Time of post-test (n=16)	
Immediate post-test (n=6)	*ES* = .093, *z* = 2.14, *p* < .05*
Post-test 1 month+ (n=10)	*ES* = .191, *z* = 5.06, *p* < .001*
Between-studies heterogeneity	Q_B = 2.93 ~χ^2_1, *p* = .087
Within-studies heterogeneity	Q_W = 39.35 ~χ^2_{14}, *p* < .05*
Sample ethnicity (n=13)[c]	
Caucasian/mixed (n=6)	*ES* = .328, *z* = 6.92, *p* < .001*
Predominant minority (n=7)	*ES* = .085, *z* = 2.00, *p* < .05*
Between-studies heterogeneity	Q_B = 25.24 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 17.04 ~χ^2_{11}, *p* = .807

a. Two studies were missing from this analysis as information on whether the program addressed gender roles was not provided.
b. Two studies were missing from this analysis as information regarding the inclusion of bystander training was not provided.
c. Three studies were missing from this analysis as information regarding sample ethnicity was not reported.

Across all seven moderating variables (Table 4.18) only three were found to be significant, with both program characteristics and only one of the study characteristics resulting in a significant *Q*-between statistic. Whether the program included information regarding gender roles was a significant moderator (Q_B = 7.88, *p* < .01); when examining

the specific effect sizes, we see programs that do not include gender role education produce significantly larger effects than those that do address gender roles and stereotypes.

Additionally, the incorporation of bystander training was a significant moderating variable for the dating violence perpetration outcome. The *Q*-between result was significant at 13.7 ($p < .001$); the effect size for programs without bystander training was significant (*ES* = .231, $z = 5.83$, $p < .001$) and significantly larger than the nonsignificant effect size produced by the bystander training group.

Of the five study characteristics, only sample ethnicity was significant, with a significant *Q*-between statistic of 25.24 ($p < .001$). For this outcome both groups had a significant effect size, though the predominantly Caucasian or mixed ethnicity group was significantly larger (*ES* = .328, $z = 6.92$, $p < .001$) compared to the studies with predominantly ethnic minority samples (*ES* = .085, $z = 2.00$, $p < .05$). This difference is consistent with the significant *Q*-between statistic and illustrates the importance of developing content specific to a target population.

The remaining study characteristics of treatment sample size, research design, publication year, and time of post-test were all insignificant as moderating variables. Although all four variables produced statistically significant effect sizes across both groups within the variable, there were no significant *Q*-between statistics. This suggests that effect sizes within each of these variables are relatively homogenous, and these variables have no impact at moderating the programs' impact on dating violence perpetration.

4.5. Outcome 4: Dating violence victimization experiences

4.5.1. Pooled analysis

Outcome measures of incidents of dating violence victimization (n=12) were pooled using a random effects model, which yielded an effect size of 0.103 ($z = 1.53$, $p = 0.125$). This result is not significant, which suggests that dating violence prevention programs overall do not have an impact on incidents of dating violence victimization among adolescents. This model also resulted in a statistically significant *Q*-statistic suggesting high levels of heterogeneity in the sample (46.53, $df = 11$, $p < .001$). The

accompanying I_2 statistic of 76.4% indicates that a majority of the heterogeneity is due to factors other than sampling error.

Table 4-19. Random effects meta-analysis for dating violence victimization experiences

Pooled estimate	95% CI Lower	95% CI Upper	Z (*p*-value)	No. of Studies	*Q*-Statistic	I_2
0.103	-0.029	0.235	1.53 (*p*=.125)	12	46.53, df=11, $p < .001$	76.4%

Table 4-20. Study level data for the meta-analysis of the dating violence victimization experiences outcome

Author, year	Effect Size	95% CI Lower	95% CI Upper	Relative weight (%)
Sanchez-Jiminez (2018)	-0.022	-0.149	0.105	11.47
Reidy (2017)	-0.059	-0.189	0.072	11.40
Leveseque (2016)	0.419	0.270	0.569	11.02
Ball (2012)	0.032	-0.130	0.194	10.75
Jaycox (2006)	-0.020	-0.192	0.152	10.54
Edwards (2019)	0.073	-0.139	0.285	9.64
Peskin (2014)	0.288	0.047	0.528	9.00
Connolly (2015)	0.655	0.291	1.019	6.50
Cissner (2014)	-0.160	-0.551	0.232	6.03
Muñoz-Fernández	-0.091	-0.498	0.316	5.78
Rizzo (2019)	-0.301	-0.831	0.229	4.19
Gonzalez-Guarda (2015)	0.456	-0.124	1.035	3.69

The forest plot in Figure 4.10 illustrates the effect of dating violence prevention programs on incidents of dating violence victimization and shows the positive, yet non-significant effect. Of the 12 included studies, six fell to the left of the zero point with negative effects, with the remaining six being positive. However, only three studies produced statistically significant effects (Levesque, 2016, Peskin, 2014, and Connolly, 2015), and three studies produced effect sizes smaller than 0.05 (Sanchez-Jiminez, 2018, Ball, 2012, and Jaycox, 2012).

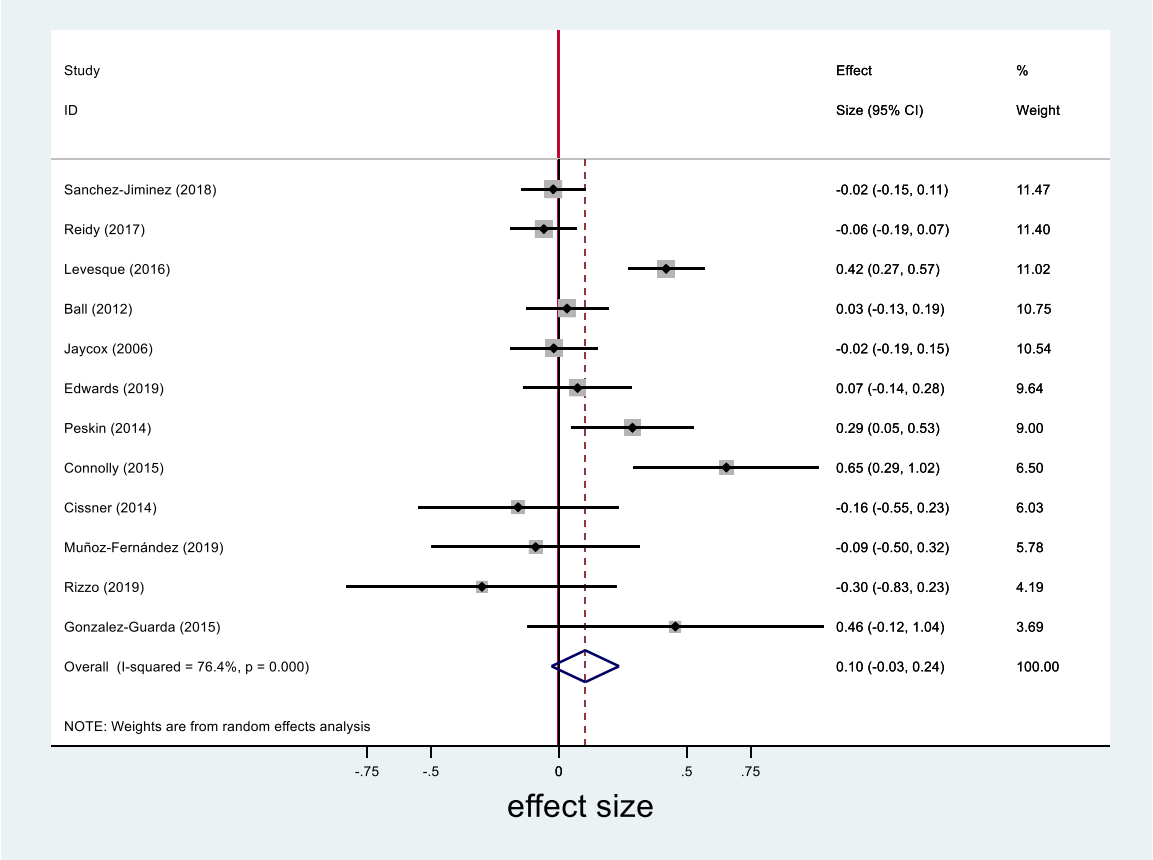

Figure 4-10. Forest plot for the violence victimization outcome

4.5.2. Sensitivity analysis

Publication bias

To determine whether publication bias was present in the dating violence victimization sample of studies, a funnel plot was visually inspected and Egger's test of small study effects was performed. The funnel plot appears to be somewhat symmetrical, with individual estimates dispersed in all areas of the funnel. However, two effects fall outside the pseudo 95% confidence interval, which warrant additional investigation.

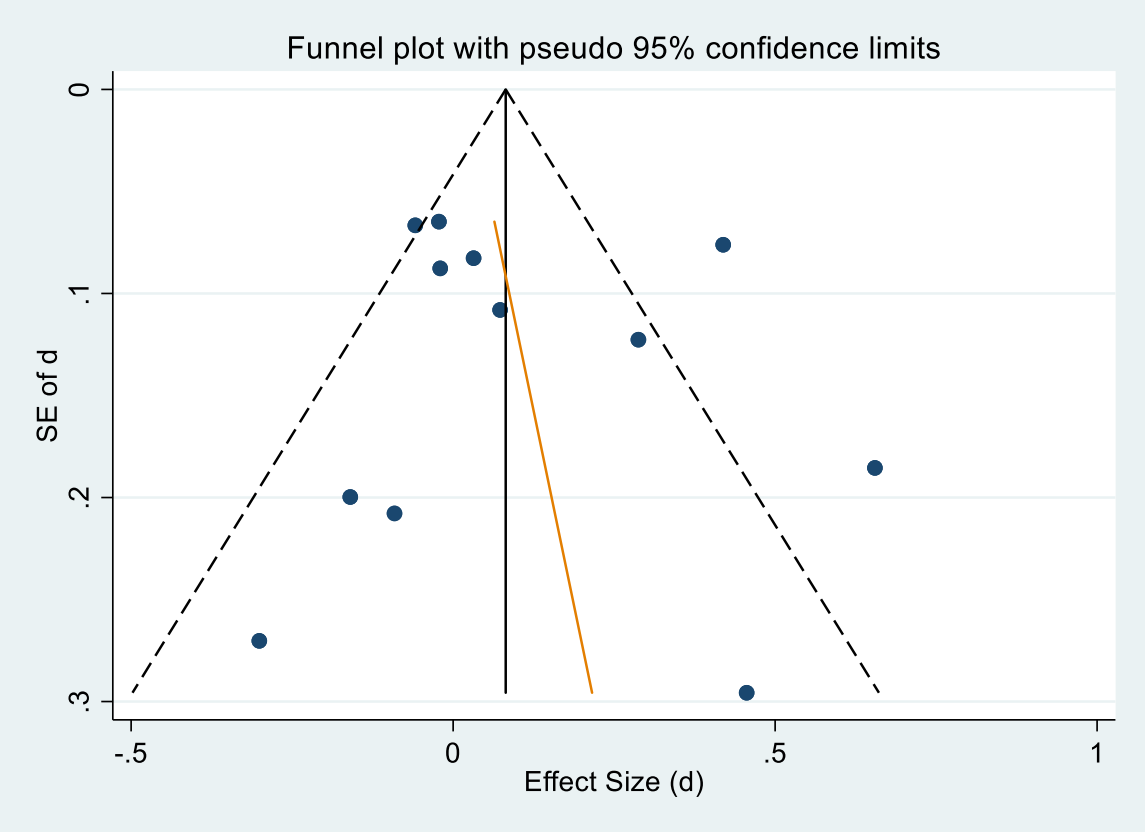

Figure 4-11. Funnel plot for the victimization outcome

Egger's test of small study effects was performed for this sample and the results are consistent with the appearance of the funnel plot. The bias coefficient was not significant (0.657, *SE* = 1.43, *t* = 0.46, *p* = 0.656), which indicates that small study effects are not a problem within the dating violence victimization sample of studies.

Influence analysis

Influence analysis was used to assess the individual impact each study had on the overall pooled effect. Table 4.21 and Figure 4.12 display the adjusted pooled effect for each study if they were removed from the analysis. The pooled effect for victimization outcomes was not statistically significant, and the removal of any single study did not change this result. Though the removal of studies such as Levesque (2016) or Connolly (2015) resulted in a change in the estimate, neither changed the level of significance of the pooled effect and thus were not determined to be exerting a concerning influence on the overall effect size.

Table 4-21. Influence analysis for experiences of dating violence

Study Omitted	Effect Size	95% CI Lower	95% CI Upper
Ball (2012)	0.112	-0.037	0.261
Cissner (2014)	0.120	-0.017	0.257
Connolly (2015)	0.066	-0.058	0.191
Edwards (2019)	0.107	-0.038	0.252
Gonzalez-Guarda (2015)	0.090	-0.045	0.224
Jaycox (2006)	0.118	-0.028	0.264
Levesque (2016)	0.055	-0.053	0.164
Munoz-Fernandez (2019)	0.115	-0.022	0.253
Peskin (2014)	0.085	-0.054	0.224
Reidy (2017)	0.124	-0.021	0.269
Rizzo (2019)	0.121	-0.013	0.255
Sanchez-Jiminez (2018)	0.119	-0.030	0.269
Combined	**0.103**	**-0.029**	**0.235**

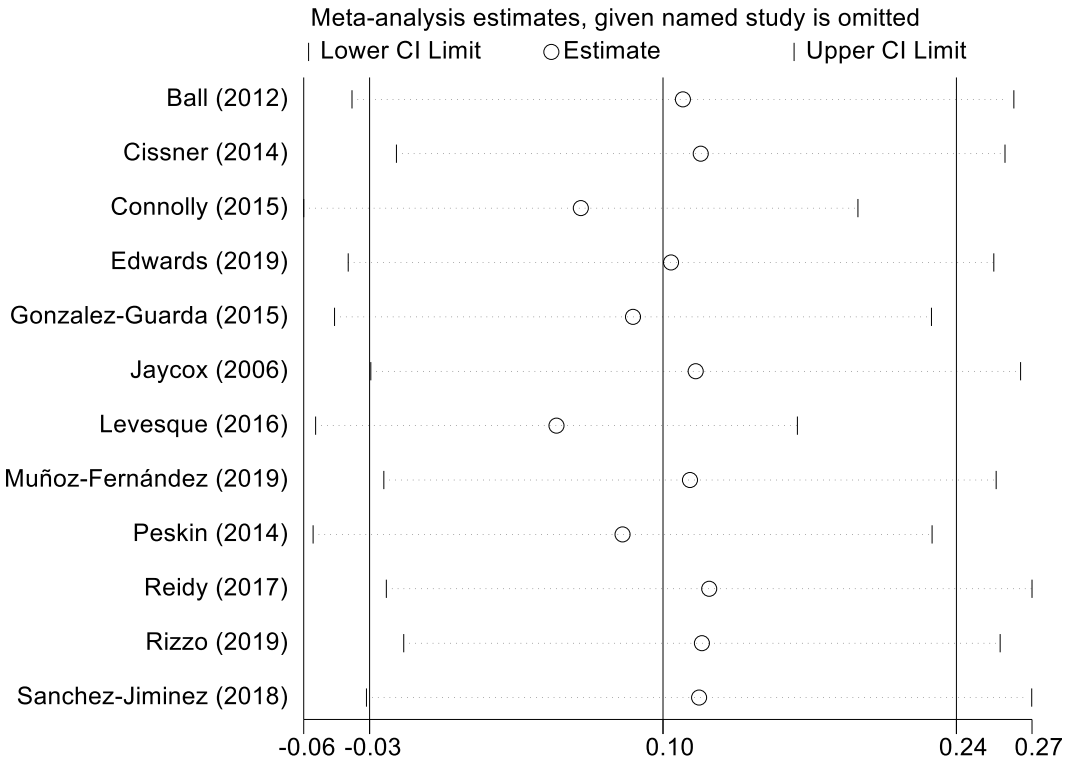

Figure 4-12. Influence analysis for incidents of dating violence victimization

4.5.3. Assessing heterogeneity

Heterogeneity between studies pooled in the dating violence victimization outcome was assessed by conducting subgroup analysis using seven dichotomous variables representing program and study characteristics: length of program (≤ 5 hours versus ≥ 6 hours); whether the curriculum included bystander training (yes/no); whether the curriculum addressed gender roles (yes/no); level of school (middle school versus high school); treatment group sample size (<300 versus ≥300); publication year (prior to 2015 versus 2015 or later); time of post-test (immediately after program end versus 1+ months after program end).

Table 4-22. Fixed effects moderator analysis for incidents of dating violence victimization

Study characteristic	Effect size and *Q*-statistics
Length of program (hours) (n=12)	
5 or fewer hours (n=4)	*ES* = .245, *z* = 4.56, *p* < .001*
6 or more (n=8)	*ES* = .013, *z* = 0.37, *p* = .710
Between-studies heterogeneity	Q_B = 13.12 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 33.41 ~χ^2_{10}, *p* = .074
Program addressed gender roles (n=12)	
No (n=8)	*ES* = .135, *z* = 3.73, *p* < .001*
Yes (n=4)	*ES* = -.018, *z* = 0.37, *p* = .715
Between-studies heterogeneity	Q_B = 6.25 ~χ^2_1, *p* < .05*
Within-studies heterogeneity	Q_W = 40.28 ~χ^2_{10}, *p* < .05*
Bystander (n=10)[a]	
No (n=5)	*ES* = .118, *z* = 2.90, *p* < .01*
Yes (n=5)	*ES* = 0.00, *z* = 0.01, *p* = .995
Between-studies heterogeneity	Q_B = 13.24 ~χ^2_1, *p* < .001*
Within-studies heterogeneity	Q_W = 33.29 ~χ^2_8, *p* = .076
Level of school (n=12)	
Middle school (n=4)	*ES* = .224, *z* = 2.59, *p* < .05*
High school (n=8)	*ES* = .063, *z* = 2.03, *p* < .05*
Between-studies heterogeneity	Q_B = 3.06 ~χ^2_1, *p* = .080
Within-studies heterogeneity	Q_W = 43.47~χ^2_{10}, *p* < .01*
Treatment group sample size (n=12)	
Less than 300 (n=4)	*ES* = .124, *z* = 1.75, *p* = .080
300 or more (n=8)	*ES* = .073, *z* = 2.27, *p* < .05*
Between-studies heterogeneity	Q_B = 0.42 ~χ^2_1, *p* = .517
Within-studies heterogeneity	Q_W = 46.11 ~χ^2_{10}, *p* < .01*
Publication year (n=12)	
Prior to 2015 (n=4)	*ES* = .047, *z* = 0.90, *p* = .370
2015 - 2019 (n=8)	*ES* = .098, *z* = 2.77, *p* < .01*
Between-studies heterogeneity	Q_B = 0.65 ~χ^2_1, *p* = .420
Within-studies heterogeneity	Q_W = 45.88 ~χ^2_{10}, *p* < .01*
Time of post-test (n=12)	
Immediate post-test (n=4)	*ES* = .004, *z* = 0.09, *p* = .930
Post-test 1 month+ (n=8)	*ES* = .146, *z* = 3.69, *p* < .001*
Between-studies heterogeneity	Q_B = 5.80 ~χ^2_1, *p* < .05*
Within-studies heterogeneity	Q_W = 40.73 ~χ^2_{10}, *p* < .05*

a. Two studies were missing from this analysis as information regarding the inclusion of bystander training was not provided.

As per Table 4.22, a significant moderating effect was found for three of the program characteristics and one of the study characteristics. First, the moderating effect of program length is statistically significant with a *Q*-between statistic of 13.12 ($p < .001$). Examination of the effect sizes for each subgroup suggest that shorter programs produced significantly larger effect sizes than longer programs. Additionally, the effect size for the shorter program subgroup was significant (*ES* = .245, *z* = 4.56, $p < .001$),

while the effect size for longer programs was not significant (*ES* = .013, *z* = .37, *p* = .710).

The inclusion of gender role education was also a significant moderating variable, which resulted in a significant *Q*-between statistic (6.25, $p < .05$). The pooled effects sizes for each subgroup indicate that programs that do not include the topic of gender roles or stereotypes in the curriculum produce significantly larger, and positive effect sizes, while those programs that do discuss gender roles produce negative (and non-significant) effects.

With respect to the inclusion of bystander training there is a similar pattern of results, with a significant *Q*-between statistic for the two subgroups (13.24, $p < .001$). The specific subgroup effects sizes suggest that programs without bystander training produce significantly larger and positive effect sizes, while those that do include bystander training produced null effect sizes.

The school level at which the program was implemented (middle versus high school) was not a significant moderator. The pooled effect size for each of the two subgroups was significant, suggesting both middle and high school implementation was effective, however there were no significant differences between subgroups ($Q_B = 3.06$, $p = .080$). The lack of a significant *Q*-between result implies a certain level of homogeneity within the school level variable, suggesting that the school level of implementation is not important for moderating the treatment effect.

Last, the timing of the post-test measure was found to have a significant moderating effect ($Q_B = 5.80$, $p < .05$). Examination of the pooled effect sizes for each of the two subgroups reveals that significantly larger effect sizes are produced by studies with a post-test that occurred at least one month after the end of the program, compared to those with an immediate post-test. This result appears to be the opposite of what would be expected; typically, we would expect to see larger effects at an immediate post-test as the effect of the program is less likely to have dissipated or weakened over time.

4.6. Outcome 5: Bystander behaviours and intentions

4.6.1. Pooled analysis

Of the full sample, seven separate studies included a measure of bystander behaviour or intentions. The random effects model resulted in a significant pooled estimate of 0.499 (z = 2.23, p < 0.05), suggesting dating violence prevention programs are effective at increasing bystander behaviours. There was also a statistically significant Q-statistic of 130.19 (df = 6, p < .001) and an I^2 *of 95.4%,* suggesting a substantial amount of heterogeneity.

Table 4-23. Random effects meta-analysis for bystander behaviours and intentions

Pooled estimate	95% CI Lower	95% CI Upper	Z (*p*-value)	No. of Studies	Q-Statistic	I^2
0.499	.061	0.936	2.23 (*p<.05*)	7	130.19, df=6, *p* < .001	95.4%

Table 4-24. Study level data for the meta-analysis of the bystander behaviours and intentions outcome

Author, year	Effect Size	95% CI Lower	95% CI Upper	Relative weight (%)
Silverman (1998)	0.016	-0.066	0.097	34.37
Miller (2012)	0.182	-0.023	0.387	19.23
Sargent (2017)	0.120	-0.085	0.325	19.23
Jouriles (2019)	0.457	0.148	0.767	11.49
Miller (2015)	0.060	-0.252	0.372	11.35
Baynard (2019)	-0.067	-0.630	0.496	4.33
Plourde (2016)	13.965	11.488	16.422	2.65

The forest plot in Figure 4.13 represents the bystander behaviour and intention outcome and the significant, positive pooled effect. Six of the seven individual effect sizes are positive, however only one is statistically significant (Jouriles, 2017). The final study (Banyard, 2019) produced the sole negative effect (though non-significant). Notably, Plourde (2016) produced a substantially larger effect size compared to the other six studies; visually it can be seen on the far right edge of the forest plot, suggesting that it may be an outlier.

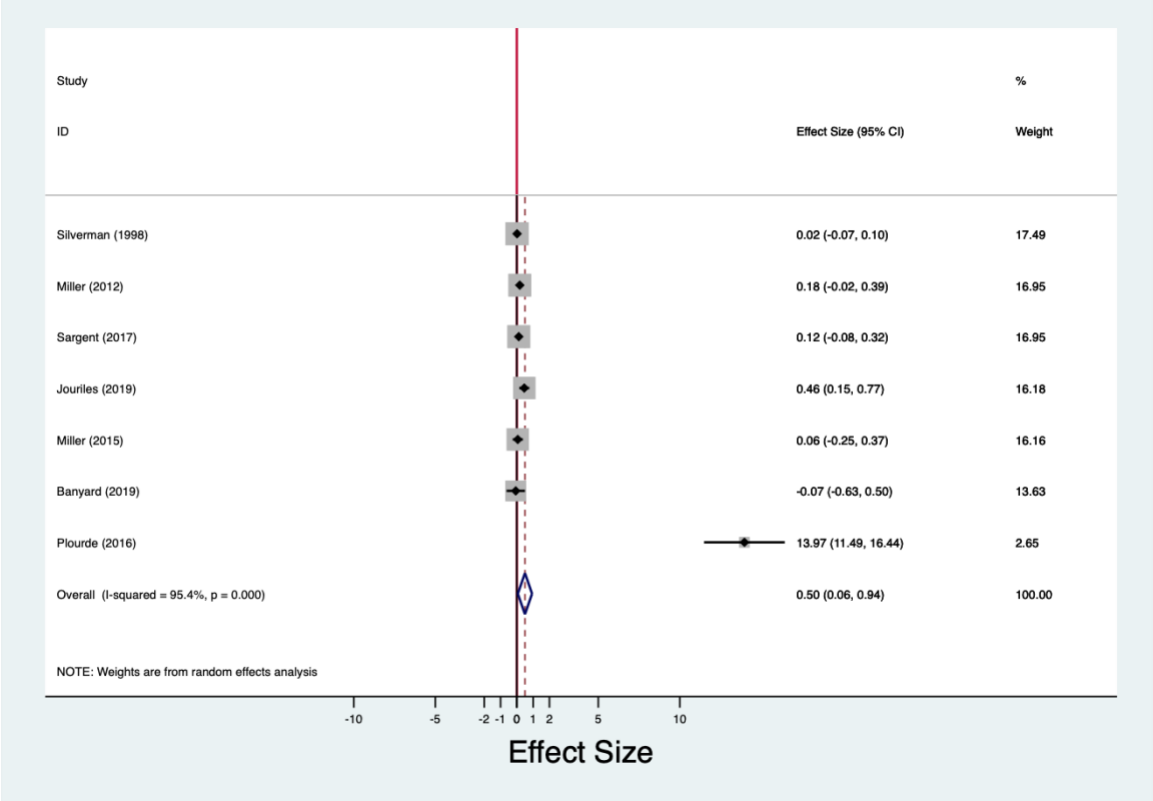

Figure 4-13. Forest plot for the bystander behaviour and intention outcome

4.6.2. Sensitivity analysis

Publication bias

Examining the funnel plot for the bystander behaviour outcome (Figure 4.14) demonstrates slight asymmetry among the estimates, with estimates missing from the lower right-hand side of the funnel. To examine the possibility of bias in more detail, Egger's test of small study effects was performed. The bias coefficient was not statistically significant (1.263, *SE* = 0.920, *t* = 1.37, *p* = 0.242), meaning there is no evidence of small study effects. However, one study was identified as a clear outlier in the plot (Plourde, 2016). As well, it is important to note that tests for funnel plot asymmetry are recommended only with a sample size of 10 studies or more to ensure adequate power (Harbord, Harris, & Sterne, 2009). The sample size for the bystander behaviour outcome is seven, and as such the results should be interpreted with caution; the results are not definitive but may indicate a problem that requires further investigation.

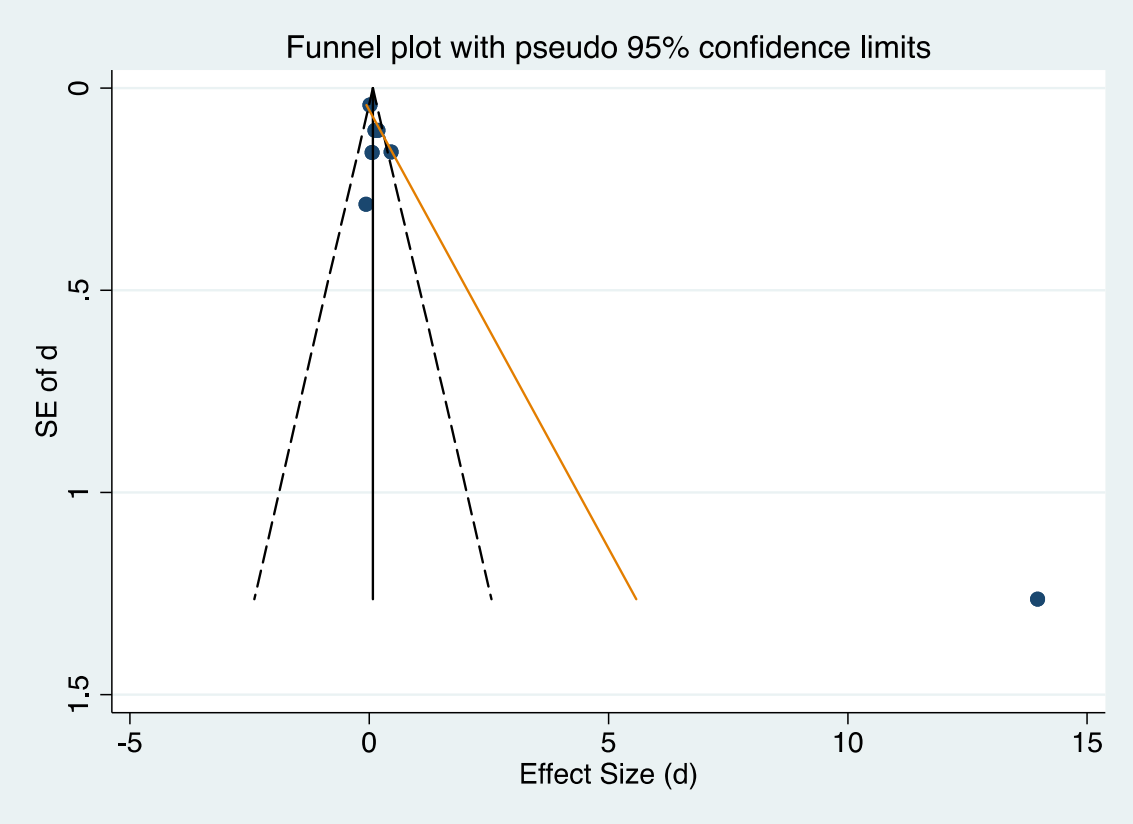

Figure 4-14. Funnel plot for bystander behaviours

Influence analysis

To assess the individual influence of each separate study on the overall pooled bystander estimate, each study was removed from the analysis, one at a time. Table 4.25 illustrates the adjusted pooled estimate with the removal of each study. Examination of the lower and upper confidence limits reveals that the removal of Plourde (2016) drops the pooled estimate significantly. Plourde (2016) was previously identified as a potential outlier due to a very large effect size (*ES* = 13.965, *SE* = 1.264). The significant drop in the pooled effect as a result of removing Plourde (2016) suggests that the result is heavily reliant on this single study, and as such is not robust. As the overall pooled effect is so reliant on the Plourde (2016) estimate, it was removed from all analyses.

Table 4-25. Influence analysis for bystander behaviours and intentions

Study Omitted	Effect Size	95% CI Lower	95% CI Upper
Banyard (2019)	0.604	0.123	1.085
Jouriles (2019)	0.541	0.040	1.041
Miller (2012)	0.691	0.128	1.254
Miller (2015)	0.634	0.123	1.145
Plourde (2016)	0.120	-0.004	0.244
Sargent (2017)	0.707	0.142	1.271
Silverman (1998)	0.890	0.222	1.559
Combined	**0.499**	**0.061**	**0.936**

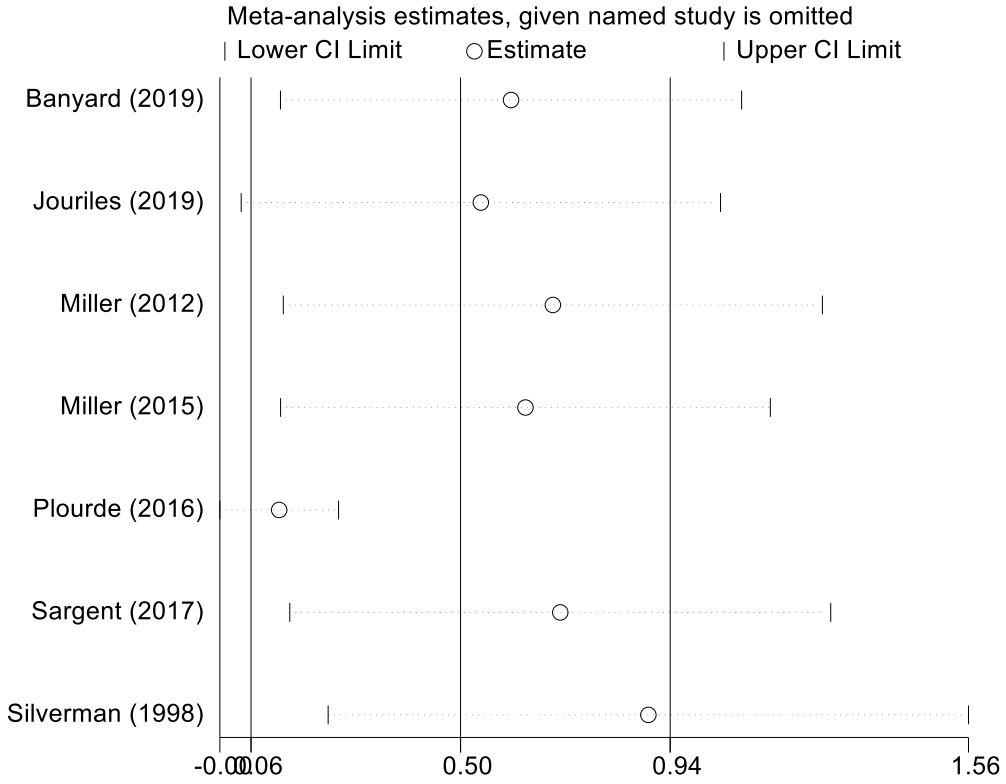

Figure 4-15. Influence analysis for bystander behaviours

After the removal of Plourde (2016), the effect sizes were pooled using a random effects model once again, which resulted in a non-significant estimate of 0.120 (z = 1.90, p = .058). This result implies that dating violence prevention programs for adolescents are not effective at significantly increasing bystander behaviours and intentions. Additionally, the Q-statistic of 9.33 (df = 5, p = .097) was not significant, suggesting the

sample is not significantly heterogeneous. This is consistent with the I_2 statistic, which suggests 46.4% of the heterogeneity can be attributed to factors other than standard error.

Table 4-26. Random effects meta-analysis for bystander behaviours and intentions (outlier removed)

Pooled estimate	95% CI Lower	95% CI Upper	Z (*p*-value)	No. of Studies	Q-Statistic	I_2
0.120	-0.004	0.244	1.90 (*p*=0.058)	6	9.33, df=5, *p* = .097	46.4%

Table 4-27. Study level data for the meta-analysis of the bystander behaviours and intentions outcome (outlier removed)

Author, year	Effect Size	95% CI Lower	95% CI Upper	Relative weight (%)
Silverman (1998)	0.016	-0.066	0.097	34.37
Miller (2012)	0.182	-0.023	0.387	19.23
Sargent (2017)	0.120	-0.085	0.325	19.23
Jouriles (2019)	0.457	0.148	0.767	11.49
Miller (2015)	0.060	-0.252	0.372	11.35
Baynard (2019)	-0.067	-0.630	0.496	4.33

The forest plot in Figure 4.16 represents the updated bystander behaviour and intention outcome and the non-significant (though positive) pooled effect, with five of the six effects being positive and one negative.

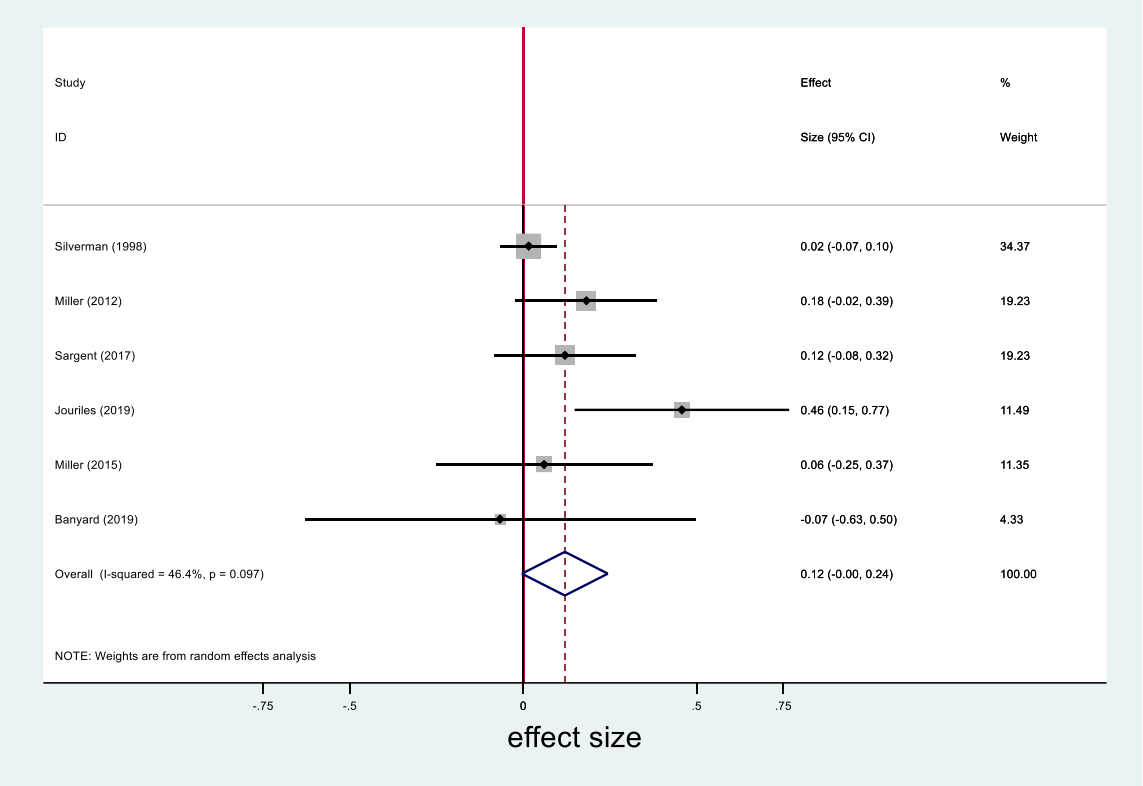

Figure 4-16. Forest plot for the bystander behaviour and intention outcome (outlier removed)

4.6.3. Assessing heterogeneity

Although the *Q*-statistic was not statistically significant, suggesting homogeneity between studies, it is still good practice to investigate whether systematic heterogeneity is present, as the *Q*-statistic is often claimed to be underpowered, and may not accurately detect heterogeneity in a smaller sample (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Unfortunately, the small sample size of six studies was too small to conduct any meaningful comparisons between subgroups.

Chapter 5.

Discussion

Dating violence refers to any acts of psychological, physical, or sexual violence that occur within the context of a dating relationship (CDC, 2019a). This issue is highly prevalent in adolescent populations, with studies reporting that up to 69% of adolescents have perpetrated dating violence or have been victimized (Hickman, et al., 2004; Taylor & Mumford, 2016). These prevalence rates are particularly concerning given that a number of adverse outcomes have been associated with dating violence; negative consequences include depression, anxiety, substance use, disordered eating, risky sexual behaviours, and physical injury (see Ackard & Neumark-Sztainer, 2002; Banyard & Cross, 2008; Callahan et al., 2003; Coker et al., 2000; Foshee et al., 2013; Silverman, et al., 2001; Silverman, et al., 2004; Tharp et al., 2017). As such, dating violence prevention programs are particularly important for teens; however, more information regarding the effectiveness of these programs is needed.

5.1. Main effects

This study examined the effectiveness of adolescent dating violence prevention programs at increasing knowledge about dating violence, improving attitudes towards and beliefs about dating violence behaviours, reducing incidents of dating violence perpetration and victimization, and increasing the prevalence of bystander behaviours and intentions. In total, the initial search identified 4,430 articles and 962 grey literature sources for review, resulting in the retrieval of 135 articles. This led to a final 37 studies included in the analysis, which produced 71 independent effect sizes. The programs were primarily delivered in a school-based, classroom setting (73%) and targeted adolescents with a mean age of 14.52 (*SD* = 1.2). Table 5.1 provides a summary of the pooled effects and, overall, dating violence prevention programs appear to be effective at improving elements of dating violence. Present evidence provides support for a positive effect for three of the five primary outcomes; however, the effect sizes are quite small for all outcomes. One exception is the pooled effect size for the knowledge outcome, which would be classified as a medium effect size as per Cohen (1988) and Lipsey and Wilson (1993). Traditionally, Cohen's guidelines are the primary estimates

used for interpreting effect size. Cohen (1988) suggests that a small effect size corresponds to 0.20 standard deviations, medium is .50 standard deviations, and large is .80 standard deviations or above. However, Lipsey and Wilson (1993) also found support for these conventions after examining effect sizes for over 300 meta-analyses of various interventions; Lipsey and Wilson found the following benchmarks: a small effect size is approximately 0.30 standard deviations, medium is 0.50 standard deviations, and a large effect size is 0.67 standard deviations or above.

Table 5-1. Summary of pooled analyses

Outcome	No. of effect sizes	Pooled estimate	Z (*p*-value)
Knowledge	16	0.566	3.59 ($p < .001$)
Attitudes towards dating violence	20	0.191	3.88 ($p < .001$)
Incidents of dating violence perpetration	16	0.156	2.94 ($p < .01$)
Experiences of dating violence victimization	12	0.103	1.53 ($p = .125$)
Bystander behaviours and intentions	6	0.120	1.90 ($p = .058$)

Dating violence prevention programs appear effective at increasing adolescents' knowledge about dating violence and its associated behaviours, with a statistically significant pooled effect size of 0.566 ($p < .001$). This result is consistent with two of the previous dating violence meta-analyses (de la Rue et al., 2017 and Ting, 2009), which also found an overall positive effect for dating violence knowledge outcomes. In addition, the pooled effect was significant and positive for changing adolescents' attitudes and beliefs towards dating violence (ES = 0.191, $z = 3.88$, $p < 0.001$), which again is consistent with previous meta-analytic literature in the area of dating violence prevention. Both de la Rue et al. (2017) and Ting (2009) found a significant improvement in dating violence attitudes. It is interesting to note that the studies included in the present study had very little overlap with those of both prior meta-analyses (six studies in common with de la Rue et al. (2017) and three studies in common with Ting (2009)). The fact that the current study found similar results as previous research regarding the outcomes of dating violence knowledge and attitudes, despite having very little overlap, suggests that teen dating violence prevention programs are consistently having a positive impact on these two outcomes. 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.

When it comes to behaviours, we see a somewhat different pattern of results. When examining incidents of dating violence perpetration (of any physical, psychological/emotional, or sexual violence), a positive impact of the prevention programs was found, with a significant decrease in perpetration behaviours among participants (ES = 0.156, $z = 2.94$, $p < 0.01$). Of the previous meta-analyses, de la Rue et al. (2017) and Fellmeth et al. (2015) examined perpetration of dating violence behaviours, however neither found a significant effect. The results of the current study may differ due to the specific studies included in the analysis; the current study had very little overlap with previous studies based on differing inclusion criteria (six studies in common with de la Rue (2017) and two studies in common with Fellmeth et al. (2015)). Previous meta-analyses included not only studies examining dating violence prevention programs, but also programs that solely addressed sexual violence (e.g., rape, acquaintance rape, sexual coercion). These results suggest that programs targeting dating violence as a whole may be effective at reducing incidents of dating violence perpetration, but programs specific to sexual violence may be less effective. However, these results may also be a function of the types of studies included and the population examined; Fellmeth et al.'s (2015) meta-analysis included studies focused on both youth and college-aged adults. Arguably, sexual violence prevention programs for college students and dating violence prevention programs for adolescents are too different to appropriately compare.

While the present study found a significant effect for dating violence perpetration, we see different results for the other two behavioural outcomes. Both dating violence victimization and bystander behaviour and intentions resulted in null pooled effects. These null effects are consistent with de la Rue et al. (2017) and Fellmeth et al. (2015), who also found null effects for measures of victimization. It is interesting to note that while significance levels were consistent, the direction of the relationship was not. Specifically, de la Rue et al. (2017) found a non-significant *negative* effect, while the present study found a non-significant *positive* effect. Additionally, while the victimization outcome was non-significant overall, there were several significant moderator variables. This suggests that dating violence prevention programs are progressing in the right direction in terms of reducing experiences of victimization; however, it is important to

examine these programs more closely to determine where the differences lie and how to better impact victimization rates.

It is also important to note that the effect on perpetration of violence was significantly positive, while the effect of victimization was not. It would be expected that both outcomes would mirror the other; if perpetration is significantly decreasing, then so should victimization (i.e., if adolescents are perpetrating less violence, there should be fewer adolescents being victimized). However, this does not appear to be the case. This finding highlights the importance of considering both perpetration and victimization as separate outcomes, with one not necessarily a direct reflection of the other. Prevention programs should consider potential factors that may influence each individual outcome.

Bystander behaviour was also not significantly impacted by the dating violence prevention programs, though the direction of the effect was generally positive. The mixed results regarding the behavioural impacts of dating violence prevention programs suggest that behaviours are more difficult to change than knowledge or beliefs. However, it is also important to consider the time of measurement for these particular outcomes. Most of the included effect sizes were based on measures taken at an immediate post-test. Knowledge and attitudes may reflect an immediate change as they do not necessarily require any implementation or practice to change. Behavioural outcomes on the other hand may take time to evidence any noticeable change. This is in part due to the lack of opportunity to implement the new knowledge in situations where these behaviours may occur. For example, if there is an immediate post-test, a participant may not yet have experienced any opportunity to engage in bystander behaviours. To fully assess the effect of dating violence prevention programs on behaviours, longer gaps between the end of programs and follow-up assessments should be examined.

5.2. Subgroup analysis

Four of the main effect analyses detected significant heterogeneity between studies. To investigate possible sources of the variability between studies, subgroup analyses were conducted using a series of nine categorical variables representing various program and study characteristics. These variables include: length of program (≤5 hours versus ≥6 hours); whether the curriculum addressed gender roles and

stereotypes (yes/no); whether the curriculum incorporated bystander training (yes/no); school level (middle school versus high school); treatment group sample size (<300 versus ≥300); research design (randomized control trial versus non-randomized); publication year (prior to 2015 versus 2015 or later); and time of post-test (immediately after program end versus 1+ months after program end). Table 5.2 provides an overview of the moderator results across the four outcomes examined.

Table 5-2. Summary of moderator analyses

Moderator	Knowledge	Attitudes towards dating violence	Incidents of dating violence perpetration	Experiences of dating violence victimization
Program characteristics				
Gender roles	✓	✓	✓	✓
Bystander training	n/a	n/a	✓	✓
Program length	✓	x	n/a	✓
School level	n/a	x	n/a	x
Study characteristics				
Timing of post-test	✓	x	x	✓
Research design	✓	✓	x	n/a
Sample ethnicity	✓	✓	✓	n/a
Sample size	x	✓	x	x
Publication year	✓	x	x	x

✓ = significant moderator; x = not a significant moderator; n/a = not examined

Program characteristics

Program length had differing results based on outcome, and was a significant moderator for both dating violence knowledge and victimization experiences, but was non-significant for attitudes towards dating violence[16]. Although program length had a significant moderating effect for both dating knowledge and victimization, the direction of the effect for each was opposite of the other. For the outcome of dating knowledge,

[16] Program length was not included in the moderator analyses for the perpetration outcome due to sample size.

longer programs produced significantly larger effect sizes; this result makes logical sense. It is expected that a longer program would be more comprehensive and would also provide the participants with more time to learn the material, thus leading to a larger effect size. Dating violence victimization, however, showed significantly larger effects produced by shorter programs. This is unexpected, but 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 and better learn the material when it is presented in a shorter time frame. This is particularly likely with younger adolescents who may be more likely to become bored or fatigued as the program progresses.

Whether gender roles were addressed in the program curriculum was a significant moderating variable for all four outcomes (knowledge, attitudes towards dating violence, violence perpetration, and victimization experiences), with significant differences between the pooled effects for subgroups of programs that did and did not address gender roles. However, the direction of the differences varied between the outcomes. When examining attitudes towards dating violence, programs that discuss gender roles produced larger effect sizes than programs that did not discuss them. This finding is not unexpected; given that gender roles and stereotypes are believed to perpetuate myths regarding gender-based violence (Sampert, 2010); we would expect that teaching adolescents about these topics could help to dispel some of the myths and shift their attitudes in a more positive direction.

For the remaining three outcome variables, the opposite effect was found. If a program did not address gender roles, the effect size was significantly larger than the effect for programs that did address gender roles. In the context of the knowledge outcome, it may be that the included focus on gender roles and stereotypes detracts from other topics regarding dating violence overall, i.e., rather than focusing on more general facts about dating violence and behaviours, attention is concentrated on gender roles.

With respect to the victimization outcome, there was a significant difference between subgroups, with programs that do address gender roles producing significantly different and negative effect sizes compared to those that do not address gender roles. The negative trend in effect for programs that discuss gender roles may be due to

adolescents developing an increased ability to recognize instances of dating violence victimization. If programs that include the topic of gender roles and stereotypes within the curriculum are in fact more comprehensive and are having a positive impact on overall knowledge of dating violence, participants may be more likely to report higher levels of victimization at post-test. They may now be able to identify instances that they did not classify as dating violence at the time of pretest, thus resulting in what appears to be a negative program effect.

A second program characteristic variable found to be a significant moderator is bystander training. Whether a program included bystander training resulted in significant differences between subgroups for both the violence perpetration and victimization outcomes, with the inclusion of bystander training producing smaller pooled effects for the perpetration outcome and null effects for the victimization outcome. This again goes against expectation; a comprehensive program that addresses both helping others and helping oneself in dating violence situations would be expected to result in more positive effects. Instead, both program characteristics produced the opposite finding. In the context of victimization, these results suggest that specific bystander training may not have an impact on participants' own experiences of victimization. As bystander training is focused on intervening in situations involving other people and diffusing the situation, these techniques may not necessarily apply when the individual is directly involved in a situation as the target of the abuse. With respect to perpetration, it may be that those who are already perpetrating violence are less likely to be influenced by learning how to help others, and thus less likely to change their behaviour.

As well, the results may be an effect of small subgroup sample sizes, as the sample size was quite small for at least one of the subgroups in each outcome/subgroup analysis due to limited data. It may be that the analysis did not have sufficient power to adequately detect the true effect. Related to this, only two program characteristics were examined due to a lack of data. Many of the studies were missing information regarding program components, which limited the number of characteristics available for use in subgroup analyses. Given that the level of program effectiveness is likely directly related to the specific content and components within a given program, future research should further investigate the impact of individual components.

Study characteristics

Multiple study characteristics were examined for their potential moderating effects, with several resulting in significant subgroup differences. The first variable is the time of post-test; this was a significant moderating variable for the outcomes of dating violence knowledge and experiences of victimization (it was not a significant moderator for attitudes towards violence or violence perpetration). For the dating violence knowledge outcome, studies with an immediate post-test produced significantly larger effects 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.

Conversely, in the context of dating violence victimization, the opposite pattern was found, with studies with an immediate post-test producing significantly smaller effect sizes compared to those with a longer follow-up period. This goes against the expected outcome. However, as discussed previously with respect to bystander behaviours, in the context of violence victimization an immediate post-test may be unlikely to capture substantial change in victimization frequency or experiences. The immediate post-test often took place during the last session of the program or within 1-2 weeks of the program ending. This short time frame allows for little opportunity for participants to experience any meaningful change in victimization. A longer follow-up period, on the other hand, would allow for possible new opportunities for victimization, and would also provide more time for the participants to notice and observe change.

An additional study characteristic that has a significant moderating effect on the knowledge and attitude outcomes is research design. This variable was categorized as randomized control trials and non-randomized designs. For both outcomes, studies with a randomized design produced significantly smaller effect sizes than studies with non-randomized designs (including single group pretest-post-test designs). This finding was not unexpected as it is well known that less rigorous research designs are more likely to overestimate the treatment effect (Card, 2011; Carlson & Schmidt, 1999; Lipsey & Wilson, 2001; Morris & DeShon, 2002). However, this finding is particularly important given the high number of included studies using a single group pretest-post-test design. While there are significant differences between the effect sizes for randomized and non-randomized designs, for both outcomes both types of research designs were still

significant and produced a positive effect. As well, for the violence perpetration outcome, the two research design categories were not significantly different, but both did produce significant effect sizes. This suggests that while non-randomized designs do produce larger effects, evaluations using rigorous randomized designs are also producing significant effects. This result implies that the treatment effect is robust enough to be detected by both types of designs.

The ethnic composition of the sample also had a significant moderating effect for dating violence knowledge, attitudes, and incidents of perpetrated violence; each effect resulted in a significant *Q*-between statistic indicating significant differences between subgroups. The effect was consistent across all three outcomes, with predominantly Caucasian (or mixed ethnicity) samples producing significantly larger effects sizes than samples predominantly composed of an ethnic minority. This is consistent with research that has found domestic violence programs are often developed for a general population, rather than a culturally specific one (Babcock et al., 2016). These results suggest that the target population and program selected for implementation should be culturally congruent. It may be that culturally-specific programs are lacking in the field of dating violence prevention, and this factor should be considered by program developers depending on the ethnic composition of the population being targeted.

Chapter 6.

Limitations

This study was designed to be comprehensive, reliable, and valid, however, there are a few limitations. These limitations include: 1) missing data; 2) possible unidentified studies; 3) potentially flawed studies; 4) exclusion of sexual violence specific programs; 5) prioritization of immediate post-test over delayed follow-up; 6) inclusion of single group studies; 7) potentially inappropriate ICC estimates; and 8) small sample size in the subgroup analyses. Each of these limitations are discussed in detail in the following section.

1) Missing data

The first limitation is the issue of missing data; missing data present numerous challenges during the analysis process. Due to inconsistent reporting across studies, a number of variables were missing complete data. For example, all of the variables relating to the program characteristics or components were missing data due to a lack of information presented or vagueness in the description. The variables included program components such as addressing gender roles, incorporating bystander training, utilizing group discussion or group activities, engaging in role-play scenarios, and focusing on skill development. However, it was often unclear as to whether the program definitely did not include the component or if the study's description simply did not mention the component, thus resulting in missing data. As there was a considerable amount of missing data within these variables, many were unusable in the subgroup analyses. Given that the effectiveness of a dating violence prevention program is largely dependent on the information presented to participants and on the teaching methods used, this information is important to consider (Nation et al., 2003).

Additionally, missing data limited the number of studies eligible for inclusion. Many studies were missing information pertinent to calculating effect sizes. When possible, study authors were contacted in an attempt to obtain the necessary data, however, this effort resulted in no usable information. As such, when the necessary information could not be inferred the study was excluded, which reduced the overall sample size.

2) Possible unidentified studies

While the primary goal of the search was to identify and retrieve every dating violence prevention program evaluation that currently exists, there is a possibility that some studies were not identified. Despite the comprehensive search with three constructs consisting of 28 search terms across 23 bibliographic databases, it is always possible that studies exist that have not been referenced in databases or that are archived using keywords that differ from those in the search strategy. The grey literature helps to minimize this risk; however, the possibility of publication bias still presents a challenge. There may be studies that have not been published at all due to null findings.

3) Potentially flawed studies

One of the limitations of meta-analysis in general is the quality of the included studies (Card, 2011). As meta-analysis is the synthesis of existing primary research, any methodological flaws or issues of bias that are present in the included studies will also be reflected in the results of the meta-analysis. Although the selection criteria and systematic coding process attempt to control for these potential issues, the assessment of quality is limited to the information presented in the documents. If the original source is biased or misrepresents the data in some way, this cannot necessarily always be determined.

4) Exclusion of sexual violence specific programs

One potential limitation regarding the selection criteria is the exclusion of programs specifically (and only) targeting sexual violence. Given that dating violence includes acts of sexual violence, it can be argued that programs focused on sexual violence prevention should be included, and that by excluding these programs the body of literature is not accurately represented. However, while we agree that acts of sexual violence can sometimes be classified as dating violence, these acts must be perpetrated in the context of a dating relationship to be considered as such. When considering sexual violence as a whole, these acts can be committed within a dating or romantic/intimate partner relationship context but can occur in other contexts as well (e.g., stranger, acquaintance, friend). Additionally, including targeted sexual violence programs narrows the definition of dating violence. For example, rather than focusing on psychological, physical, and sexual measures of violence, the measures would primarily

focus on sexual violence. This restrictive focus on sexual violence would not provide a comprehensive definition of dating violence. Similarly, it can be argued that the purpose and goals of a sexual violence prevention program are not commensurate with the goals of programs seeking to prevent dating violence. If a program is solely focused on preventing sexual violence, it excludes all other aspects of dating violence, including knowledge and attitudes of dating violence, and emotional and physical violence. As such, it may not be measuring the same treatment effect and should not be combined with programs targeting dating violence overall.

5) Prioritization of immediate post-test over delayed follow-up

An additional potential limitation of the selection criteria is the decision rule to prioritize an immediate post-test rather than a longer follow-up measure. Arguably, a more delayed follow-up measure would have been a more informative choice as it would offer a more conservative estimate of treatment effect. As well, behavioural impacts may be less detectable at an immediate post-test; adolescents may not have the opportunity to implement the lessons they learn or engage in behavioural changes between the program start and immediate post-test. Due to differences across study designs, we believe that prioritizing the immediate post-test maximized commensurability between effect sizes. The delayed effect is important to investigate, however, and should be examined in future research to determine if the effects of dating violence prevention programs persist over time.

6) Inclusion of single group studies

As well, this study did not restrict research design to two-group designs. The selection criteria also included single group, pre-test-post-test designs; arguably, this decision could be problematic for the pooled analysis. Generally speaking, 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 in meta-analysis (see Borenstein & Hedges, 2019; Carlson & Schmidt, 1999; Cuijpers et al., 2016; Morris & DeShon, 2002). Our inclusion 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 (i.e., determining whether all studies are measuring the same treatment effect and transforming effect sizes into the same metric). The inclusion of single group

and two-group 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.

7) Potentially inappropriate ICC estimates

There is also a possible limitation regarding the choice of ICC estimates during the cluster adjustment analyses. The literature on specific guidelines for attitudinal and behavioural outcomes in an adolescent and school population is severely lacking, with the WWC (2017) guidelines being one of the few available. However, these guidelines refer more broadly to educational outcomes and are quite conservative. To account for this, we searched the existing literature for empirically-reported ICCs of relevant outcomes. Few studies reported ICCs for the exact outcomes of the current study, meaning the ICCs chosen may not necessarily be accurate. However, a substantial amount of literature was consulted, with numerous ICCs found; by using a large range of prior literature on ICCs to develop the estimates, we believe those applied in the current study are well-informed. Additionally, many existing meta-analyses have not adjusted for cluster assignments, 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 and strengthens the methodology of the current study.

8) Small sample size in the subgroup analyses

Finally, the effects of the subgroup analyses must be interpreted with caution due to small sample sizes. It is important to note that the subgroup sample sizes for several variables and outcome measures are quite small. For example, the variable regarding the discussion of gender roles in the program had small, unequal sample sizes for all four relevant outcomes (i.e., one subgroup in each outcome category has only four studies). As such, there may not be enough power to truly detect the impact of certain moderating variables, and the effects may be confounded with other characteristics of these studies or programs. Different results may have been found were the sample sizes larger.

Chapter 7.

Conclusions

Overall, this study found mixed results concerning the effectiveness of dating violence prevention programs. While there is some consistency with past meta-analytic research on program efficacy, the current analysis includes important distinctions. This study focused exclusively on dating violence prevention programs, rather than mixing dating violence and sexual violence programs. This distinction allowed for a more targeted summary regarding the impact of dating violence prevention programs alone, as compared to a conflated combination of dating violence and other violence programs. A more targeted examination allows for a closer analysis of specific program factors which may be contributing to the impact on dating violence.

The current study found adolescent dating violence prevention programs to be generally effective at increasing knowledge about dating violence, improving attitudes towards dating violence, and reducing the number of incidents of dating violence perpetration, however, programs were less successful at decreasing dating violence victimization rates or increasing bystander intentions and/or behaviours. This discrepancy suggests that current programs are better able to address knowledge and attitudinal measures of dating violence, but have less success in impacting behavioural measures. As such, existing and future programming should consider possible factors that may be influencing knowledge and attitudinal measures versus behavioural measures.

Subgroup analysis demonstrated the importance of specific program and study characteristics. While there were significant differences in effects based on the presence of certain program components (e.g., whether the program addressed gender roles or bystander training), it is important to interpret these results with caution. There was substantial variability and uncertainty in the reporting of program characteristics, limiting what could and could not be coded. Future research should examine these characteristics in depth before drawing definitive conclusions regarding their impact on program effectiveness.

In sum, dating violence prevention programs for adolescents are effective at improving knowledge and attitudinal measures of dating violence, but are not as successful at improving related behaviours. These results provide support for the continued implementation of these programs with an adolescent population but suggest there is room for improvement with respect to program approach and content. Arguably, changes in behavioural outcomes are the most pertinent to the prevention of dating violence; as such it is necessary to determine how these programs can best enact positive change. Relatedly, one of the limitations of the current analysis was the lack of reported data in the primary evaluation studies. Researchers should ensure that comprehensive reporting methods are employed to allow for a detailed understanding of dating violence prevention programs. Given that intimate partner violence is a problem that often begins in adolescence, it is imperative that programs designed to address it early and prevent its persistence are in fact effective at achieving this goal.

References

*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.

Anderson, L. A., & Whiston, S. C. (2005). Sexual assault education programs: A meta-analytic examination of their effectiveness. *Psychology of Women Quarterly, 29*, 374-388.

*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: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.

Babcock, J., Armenti, N., Cannon, C., Lauve-Moon, K., Buttell, F., Ferreira, R., . . . Solano, I. (2016). Domestic violence perpetrator programs: A proposal for evidence-based standards in the United States. *Partner Abuse*, 7(4), 355-460.

Baker, C. K. & Carreno, P. (2016). Understanding the role of technology in adolescent dating and dating violence*. Journal of Children and Familly Studies, 25*, 308-320. doi: 10.1007/s10826-015-0196-5

*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:10.1177/1077801212455188

Banyard, V. L., & Cross, C. (2008). Consequences of teen dating violence: Understanding interevening variables in ecological context. *Violence Against Women, 14*(9), 998-1013. doi: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: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: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. doi:10.1093/cs/cdx007

Bonomi, A. E., Anderson, M. L., Nemeth, J., Bartle-Haring, S., Buettner, C., & Schipper, D. (2012). Dating violence victimization across the teen years: Abuse frequency, number of abusive partners, and age at first occurrence. *BMC Public Health, 12*(637), 1-10.

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 syntehsis and meta-analysis* (3rd edition. pp. 207-243). New York: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Introduction to meta-analysis. Chichester, U.K.: John Wiley & Sons.

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.

Break the Cycle. (2010). State law report cards. Retrieved from: https://www.breakthecycle.org/sites/default/files/pdf/2010-Dating-Violence-State-Law-Report-Card-Full-Report.pdf.

Brecklin, L. R., & Forde, D. R. (2001). A meta-analysis of rape education programs. *Violence and Victims, 16*(3), 303-321.

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

Bushman, B. J. & Wang, M. C. (2009). Vote-counting procedures in meta-analysis In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research syntehsis and meta-analysis* (2rd edition. pp. 207-220). New York: Russell Sage Foundation.

Callahan, M. R., Tolman, R. M., & Saunders, D. G. (2016). Adolescent dating violence victimization and psychological well-being. *Journal of Adolescent Research, 18*(6), 664-681. doi:10.1177/0743558403254784

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

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental deisgn on effect size: Findings from the research literature on training. *Journal of applied psychology, 84*(6), 851-862.

Cascardi, M., Brown, C., Iannarone, M., & Cardona, N. (2014). The problem with overly broad definitions of bullying: Implications for the schoolhouse, the statehouse, and the ivory tower. *Journal of School Violence, 13*, 253-276. doi: 10.1080/15388220.2013.846861

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. doi:10.1177/0886260518798357

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

Centers for Disease Control and Prevention (2019b). Preventing sexual violence. Retrieved from: https://www.cdc.gov/violenceprevention/pdf/SV-Factsheet.pdf

Centers for Disease Control and Prevention (n.d.). Technology and youth: Protecting your child from electronic aggression. Retrieved from: https://www.cdc.gov/violenceprevention/pdf/EA-TipSheet-a.pdf

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

Chase, K. A., Treboux, D., & O'Leary, D. K. (2002). Characteristics of high risk adolescents' dating violence. *Journal of Interpersonal Violence, 17*(1), 33-49

*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

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

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

Connolly, J., Furman, W., & Konarski, R. (2000). The role of peers in the emergence of heterosexual romantic relationships in adolescence. *Child Development, 71*(5). 1395-1408.

*Connolly, J., Josephson, W., Schnoll, J., Simkins-Strong, E., Pepler, D., MacPherson, A., . . . Jiang, D. (2014). 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. doi:10.1177/0272431614535090

Cornelius, T. L., & Resseguie, N. (2007). Primary and secondary prevention programs for dating violence: A review of the literature. *Aggression and Violent Behavior, 12*(3), 364-375. doi:10.1016/j.avb.2006.09.006

Cornell, D. & Limber, S. P. (2015). Law and policy on the concept of bullying at school. *American Psychologist, 70*(4), 333-343. doi: 10.1037/a0038558

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. doi:10.1177/1077801218815778

Crowell, N. A., & Burgess, A. W. (Eds.). (1996). Understanding violence against women. Washington, DC: National Academy.

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. doi:10.1017/S2045796016000809

Cutbush, S., Williams, J., Miller, S., Gibbs, D., & Clinton-Sherrod, M. (2018). Longitudinal patterns of electronic teen dating violence among middle school students. *Journal of Interpersonal Violence*, 1-21. doi:10.1177/0886260518758326

*de Graaf, I., de Haas, S., Zaagsma, M., & Wijsen, C. (2015). Effects of Rock and Water: an intervention to prevent sexual aggression. *Journal of Sexual Aggression, 22*(1), 4-19. doi:10.1080/13552600.2015.1023375

De La Rue, L., Polanin, J. R., Espelage, D. L., & Pigott, T. D. (2016). 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. doi:10.3102/0034654316632061

*DeGannes, A. B. (2009). *Evaluation of sexual assualt and daating violence prevention program for middle school students* (Doctor dissertation). Temple University.

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. doi: 10.1016/j.adolescence.2016.10.002

Draucker, C., & Martsolf, D. (2010). The role of electronic communication technology in adolescent dating violence. *Journal of Child and Adolescent Psychiatric Nursing*, 23(3), 133-142.

*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:10.1007/s11121-019-01000-w

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. doi:10.1177/2158244014535787

Egger, M., Jüni, P., Bartlett, C., Holenstein, F., & Sterne, J. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment, 7*.

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.

Ellis, W. E. & Wolfe, D. A. (2015). Bullying predicts reported dating violence and observed qualities in adolescent dating relationships. *Journal of Interpersonal Violence, 30*(17), 3043-3064. doi: 10.1177/0886260514554428

Fairbairn, J. & Dawson, M. (2013). Canadian news coverage of intimate partner homicide: Analyzing changes over time. *Feminist Criminology, 8*(3), 147-176. DOI: 10.1177/1557085113480824

Fellmeth, G., Heffernan, C., Nurse, J., Habibula, S., & Sethi, D. (2014). Educational and skills-based interventions to prevent relationship violence in young people. *Research on Social Work Practice, 25*(1), 90-102. doi:10.1177/1049731514533392

Flores, S. A., & Hatlaub, M. G. (1998). Reducing rape myth acceptance: A meta-analysis of intervention studies. *Journal of College Student Development, 39*, 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*, *50*(5), 607 - 616. 10.1111/j.1469-7610.2008.02025.x.

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

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

Foshee, V. A., McNaughton Reyes, H. L., Vivolo-Kantor, A. M., Basile, K. C., Chang, L. Y., Faris, R., & Ennett, S. T. (2014). Bullying as a longitudinal predictor of adolescent dating violence. *Journal of Adolescent Health, 55*(3), 439-444. doi:10.1016/j.jadohealth.2014.03.004

Foshee, V. A., Reyes, H. L., Gottfredson, N. C., Chang, L. Y., & 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 Adolscent Health, 53*(6), 723-729. doi: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.

Garthe, R. C., Sullivan, T. N., & McDaniel, M. A. (2016). A meta-analytic review of peer risk factors and adolescent dating violence. *Psychology of Violence, 7*(1), 45-57. doi: 10.1037/vio0000040

Giordano, P. C., Kaufman, A., Manning, W. D., & Longmore, M. A. (2015). Teen dating violence: The influence of friendships and school context. *Sociol Focus, 48*(2), 150-171. doi:10.1080/00380237.2015.1007024

*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:10.1177/1059840515598843

Harbord, R. M., Harris, R. J., & Sterne, J. S. C. (2009). Updated tests for small study effects in meta-analyses. *The Stata Journal, 9*(2). 197-210.

Hawley, A., Black, B., Hoefer, R., & Barnett, T. M. (2017). domestic violence shelters' efforts to prevent teen dating violence: A national survey. *Violence Against Women, 23*(4), 520-529. doi:10.1177/1077801216644073

Hedges, L. V. (2000). Meta-analysis. In *Encyclopedia of psychology, Vol. 5.* (pp. 202-204).

Hedges, L. & Rhoads, C. (2009). *Statistical power analysis in education research* (NCSER 2010-3006). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

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. doi:10.1177/1524838003262332

Hoefer, R., Black, B., & Ricard, M. (2015). The impact of state policy on teen dating violence prevalence. *J Adolesc, 44*, 88-96. doi: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.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11*(2), 193-206. doi:10.1037/1082-989X.11.2.193

Hydén, M. (1995). Verbal aggresion as prehistory of woman battering. *Journal of Family Violence, 10*(1).

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.

*Jaycox, L. H., McCaffrey, D., Eiseman, B., Aronoff, J., Shelley, G. A., Collins, R. L., & Marshall, G. N. (2006). Impact of a school-based dating violence prevention program among Latino teens: randomized controlled effectiveness trial. *Journal of Adolescent Health, 39*(5), 694-704. doi: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., Krauss, A., Vu, N. L., Banyard, V. L., & McDonald, R. (2018). Bystander programs addressing sexual violence on college campuses: A systematic review and meta-analysis of program outcomes and delivery methods. *Journal of American College Health, 66*(6), 457-466. doi:10.1080/07448481.2018.1431906

*Jouriles, E. N., McDonald, R., Rosenfield, D., & Sargent, K. S. (2019). Increasing bystander behavior to prevent adolescent relationship violence: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 87*(1), 3-15. doi:10.1037/ccp0000355

Kann, L., McManus, T., Harris, W. A., Shanklin, S. L., Flint, K. H., Queen, B., … Ethier, K. A., (2018). Youth Risk Behavior Surveiillance – United States, 2017. Morbidity and *Mortaility Weekly Report Surveillance Summaries, 67*(8), 1-113.

Kerig, P. K. (2010). Adolescent dating violence in context: Introduction and overview. *Journal of Aggression, Maltreatment & Trauma, 19*(5), 465-468. doi:10.1080/10926771.2010.495033

Konstantopolous, S. & Hedges, L. V. (2019). Statistically analyzing effect sizes: Fixed- and random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research syntehsis and meta-analysis* (3rd edition. 245-279). New York: Russell Sage Foundation.

Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A. M. K., Hammerstrøm, K., & Sathe, N. (2017). Searching for studies: a guide to information retrieval for Campbell systematic reviews. *Campbell Systematic Reviews, 13*(1), 1-73. doi:10.4073/cmg.2016.1

Lapierre, A., Paradis, A., Todorov, E., Blais, M., & Hebert, M. (2019). Trajectories of psychological dating violence perpetration in adolescence. *Child Abuse and Neglect, 97*, 104167. doi:10.1016/j.chiabu.2019.104167

Lasserson T.J., Thomas J., Higgins J. P. T., (2019). Starting a review. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds). *Cochrane Handbook for Systematic Reviews of Interventions version 6.0.* Retrieved from: https://training.cochrane.org/handbook/current/chapter-01

Latzman, N. E., Vivolo-Kantor, A. M., Holditch Niolon, P., & Ghazarian, S. R. (2015). Predicting adolescent dating violence perpetration: role of exposure to intimate partner violence and parenting practices. *American Journal of Preventitive Medicine, 49*(3), 476-482. doi:10.1016/j.amepre.2015.06.006

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. doi:10.1016/j.avb.2012.11.015

Lehrer, J. A., Buka, S., Gortmaker, S., & Shrier, L. A. (2006). Depressive symptomatology as a predictor of exposure to intimate partner violence among US female adolescents. *Archives and Pediatrics and Adolescent Medicine, 160*, 270-276.

*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:10.1037/vio0000049

Lichty, L. F. & Gowen, L. K. (2018). Youth response to rape: Rape myths and social support. *Journal of Interpersonal Violence,* 1-28. doi: 10.1177/0886260518805777

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioural treatment: Confirmation from meta-analysis. *American Psychologist, 48,* 1181-1209.

Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Applied Social Research Methods Series: Volume 49. Thousand Oaks, CA: 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:10.1300/J202v06n03_05

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

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-282.

*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. doi:10.1093/cs/cdv006

Meilstrup, C., Ersboll, A. K., Nielsen, L., Koushede, V., Bendtsen, P., Due, P., & Holstein, B. E. (2015). Emotional symptoms among adolescents: Epidemiological analysis of individual-, classroom- and school-level factors. *European journal of Public Health, 25*(4), 644-649. doi: 10.1093/eurpub/ckv046

*Miller, E., Goldstein, S., McCauley, H. L., Jones, K. A., Dick, R., N., Jetton, J., . . . Tancredi, D. J. (2015). A school health center intervention for abusive adolescent

relationships: A cluster RCT. *Pediatics, 135*(1), 76-85. doi:10.1542/peds.2014-2471

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

*Miller, J. R. (1998). *A social learning perspective toward the prevention of dating violence: An evaluation of a group counseling model.* (Doctoral dissertation). Miami Institute of Psychology of the Caribbean Center for Advanced Studies. Miami, Florida (UMI Number: 9904710).

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. doi:10.1037/1082-989x.7.1.105

*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*(3). doi: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. doi: 10.1177/0193841X02239019

Nation, M., Crusto, C., Wandersman, A., Kumpfer, K. L., Seybolt, D., Morrissey-Kane, E., & Davino, K. (2003). What works in prevention. Principles of effective prevention programs. *American Psychologist, 58*(6-7), 449-456. doi: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., . . . 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*(2 Suppl 2), S5-13. doi:10.1016/j.jadohealth.2014.07.019

Oberle, E. & Schonert-Reichl, K. A. (2016). Stress contagion in the classroom? The link between classroom teacher burnout and morning cortisol in elementary school students. *Social Science & Medicine, 159,* 30-37. doi: 10.1016/j.socscimed.2016.04.031

Orpinas, P., Nahapetyan, L., Song, X., McNicholas, C., & Reeves, P. M. (2012). Psychological dating violence perpetration and victimization: trajectories from middle to high school. *Aggressive Behaviour, 38*(6), 510-520. doi:10.1002/ab.21441

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

Pittman, A. L., Wolfe, D. A., & Wekerle, C. (2008). Strategies for evaluating dating violence prevention programs. *Journal of Aggression, Maltreatment & Trauma, 4*(1), 217-238. doi:10.1300/J146v04n01_10

*Plourde, C., Shore, N., Herrick, P., Morrill, A., Cattabriga, G., Bottino, L., . . . Stromgren, C. (2015). You the Man: Theater as bystander education in dating violence. *Arts & Health, 8*(3), 229-247. doi:10.1080/17533015.2015.1091017

Price, E. L., & Byers, E. S. (1999). The Attitudes Towards Dating Violence Scales: Development and initial validation. *Journal of Family Violence, 14*(4), 351–375.

*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. doi:10.1016/j.ypmed.2017.05.003

Reyes, H. L., Foshee, V. A., Tharp, A. T., Ennett, S. T., & Bauer, D. J. (2015). Substance use and physical dating violence: The role of contextual moderators. *American Journal of Preventitive Medicine, 49*(3), 467-475. doi:10.1016/j.amepre.2015.05.018

*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:10.1007/s11121-018-0871-z

Rotenburg, C. & Cotter, A. (2018). Police-reported sexual assaults in Canada before and after #MeToo, 2016 and 2017. *Juristat*. Statistics Canada. Retrieved from: https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2018001/article/54979-eng.pdf?st=ylg-j_4m

Ryan, K. M. (1995). Do courtship-violent men have charactersitcs associated with a "battering personality"? *Journal of Family Violence, 10*(1), 99-120

Saarento, S., Boulton, A. J., & Salmivalli, C. (2015). Reducing bullying and victimization: Student-and classroom-level mechanisms of change. *Journal of Abnormal Child Psychology, 43,* 61-67. doi: 10.1007/s10802-013-9841-x

Salmivalli, C., Voeten, M., & Poskiparta, E. (2011). Bystanders matter: Associations between reinforcing, defending, and the frequency of bullying behavior in classrooms. *Journal of Clinical Child & Adolescent Psychology, 40*(5), 668-676. doi: 10.1080/15374416.2011.597090

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. doi: 10.3138/cjwl.22.2.301.

*Sanchez-Cesareo, M. (2002). *Outcome evaluation of the youth project: A school based teen dating violence prevention program.* (Doctoral dissertation). DePaul University, Chicago, Illinois. (UMI Number: 3076224).

*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), e0205802. doi: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. doi: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: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. doi:10.1080/01494929.2012.714722

*Silverman, J. G. (1998). *An outcome evaluation of a school-based curriculum for the pirmary preventino of dating violence among adolescents.* (Doctora dissertation). The Univeristy of Georgia, Athens, Georgia

Silverman, J. G., Raj, A., & Clements, K. (2004). Dating violence and associated sexual risk and pregnancy among adolescent girls in the united states. *Pediatrics, 114*(2), e220-e225. doi:10.1542/peds.114.2.e220

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, Pediatric, 572-579*(5), 572-.

Shapiro, J., Baumeister, R., & Kessler, J. (1991). A three-component model of children's teasing: aggression, humor, and ambiguity. *Journal of Social and Clinical Psychology*, 10(4), 459-472.

Shorey, R. C., Cohen, J. R., Lu, Y., Fite, P. J., Stuart, G. L., & Temple, J. R. (2017). Age of onset for physical and sexual teen dating violence perpetration: A longitudinal investigation. *Preventive Medicine, 105*, 275-279. doi:10.1016/j.ypmed.2017.10.008

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

*Southgate, K. (2016). *A Mixed Methods Exploration of the Impact of a Healthy Relationship Programme on Adolescents in the UK.* (Doctoral dissertation). University of Surrey, Guilford, Surrey, United Kindgdom.

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 Clincal Psychology, 71*(1), 3-13. doi: 10.1037/0022-006X.71.1.3

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.

Straus, M. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and Family, 41*(1), 75-88. doi:10.2307/351733

Sutton, A. J. (2009). Publication Bias In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2 ed., pp. 436-452). New York: Russell Sage Foundation.

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. doi:10.1177/0886260514564070

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). Retrieved from http://digitalcommons.library.tmc.edu/childrenatrisk

Temple, J. R., Shorey, R. C., Fite, P., Stuart, G. L., & Le, V. D. (2013). Substance use as a longitudinal predictor of the perpetration of teen dating violence. *Journal of Youth and Adolescence, 42*(4), 596-606. doi:10.1007/s10964-012-9877-1

Tharp, A. T., Reyes, H. L. M., 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, and Trauma, 26*(5), 445-461. doi: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. doi: 10.1016/j.jsp.2017.03.002

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

Tobias, A. (1999) Assessing the influence of a single study in the meta-analysis estimate. *Stata Technical Bulletin 47*, 15-17.

U.S. Department of Health and Human Services (2013). Frequently asked questions about the implications of clustering in clustered randomized controlled trials (RCTs). Retrieved from:

https://www.hhs.gov/ash/oah/sites/default/files/ash/oah/oah-initiatives/for-grantees/assets/tpp-evaluation-5.pdf

Vagi, K. J., Rothman, E. F., Latzman, N. E., Tharp, A. T., Hall, D. M., & Breiding, M. J. (2013). Beyond correlates: a review of risk and protective factors for adolescent dating violence perpetration. *Journal of Youth and Adolescence, 42*(4), 633-649. doi:10.1007/s10964-013-9907-7

Vezina, J., Hebert, M., Poulin, F., Lavoie, F., Vitaro, F., & Tremblay, R. E. (2015). History of family violence, childhood behavior problems, and adolescent high-risk behaviors as predictors of girls' repeated patterns of dating victimization in two developmental periods. *Violence Against Women, 21*(4), 435-459. doi:10.1177/1077801215570481

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

Widom, C. S. (1989). Does violence beget violence? A critical examination of the literature. *Psychological Bulletin, 106*(1), 3-28.

Wincentak, K., Connolly, J., & Card, N. (2017). Teen dating violence: A meta-analytic review of prevalence rates. *Psychology of Violence, 7*(2), 224-241. doi:10.1037/a0040194

*Wolfe, D. A., Jaffe, P., Chiado, 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.

Wolfe, D. A., Scott, K., Reitzel-Jaffe, D., Wekerle, C., Grasley, C., & Straatman, A.-L. (2001). Development and validation of the Conflict in Adolescent Dating Relationships Inventory. *Psychological Assessment, 13*(2), 277-293. doi:10.1037/1040-3590.13.2.277

Appendix A.

List of researcher curriculum-vitae searched

Kaleab Abebe
Barbara Ball
Amanda Borsky
Ann Coker
Vangie Foshee
Judith Herman
Denise Hines
Peter Jaffe
Ernest Jouriles
Francine Lavoie
Deborah Levesque
Elizabeth Miller
Mary Moynihan
Noelia Munoz-Fernandez
Kathleen Palm Reed
Sharyn Potter
Christie Rizzo
Emily Rothman
Bruce Taylor
David A. Wolfe

Appendix B.

List of relevant organizations searched

Battered Women's Support Services (BWSS)
Canadian Red Cross
Canadian Women's Foundation
Centers for Disease Control and Prevention (CDC)
Department of Justice (DOJ) – Office on Violence Against Women
Ending Violence Association of BC (EVA BC)
Far from the Heart
METRAC: Action on Violence
National Center on Domestic and Sexual Violence
National Institute of Justice (NIJ)
National Resource Center on Domestic Violence
Public Safety Canada
Rape, Abuse, & Incest National Network (RAINN)
RESOLVE Alberta
Violence Against Women Learning Network
Women Against Violence Against Women (WAVAW)