



Evaluating Mentors in Violence Prevention: A Longitudinal, Multilevel Assessment of Outcome Changes

Stefania Pagani¹ · Simon C. Hunter^{1,2,3}  · David Lawrence⁴ · Mark A. Elliott¹

Received: 7 February 2023 / Accepted: 17 April 2023 / Published online: 25 April 2023
© The Author(s) 2023

Abstract

There is a need to increase understanding of the effectiveness of bystander programmes targeting gender-based violence in the United Kingdom. There is also a need to utilise a robust theoretical models of decision-making while doing so. Changes were examined in bystanders' attitudes, beliefs, motivations towards intervening, and intervention behavior in situations of gender-based violence. To achieve this, a quantitative examination of Mentors in Violence Prevention was conducted. There were 1396 participants (50% female, 50% male) who were aged 11 to 14 years old ($M = 12.25$, $SD = 0.84$) attending high school at the first time point. Participants were attending 17 schools (53% Mentors in Violence Prevention and 47% control) in Scotland. Outcome variables were assessed approximately one year apart using questionnaires. Multilevel linear regressions revealed that Mentors in Violence Prevention did not change outcomes reflecting bystanders' attitudes, beliefs, motivations towards intervening, or intervention behavior in gender-based violence. Discrepancies between the current findings and those of other evaluations may be due to other studies including small numbers of schools that may be more motivated to implement the program. This study also identified two key issues that need to be addressed at stakeholder level before concluding that Mentors in Violence Prevention is ineffective at targeting gender-based violence. That the program has moved towards a more gender-neutral approach in the United Kingdom could explain the null results of this study. Furthermore, the current findings could be attributed to a failure to adequately address the theoretical model underpinning the program in practice.

Keywords Program evaluation · High school intervention · Gender-based violence · Adolescents · Bystander intervention

Introduction

Gender-based violence is violence aimed at someone because of the gender with which they identify (Morrison et al., 2007). It is one of the most common types of violence and it disproportionately targets females (Ellsberg et al., 2020). Research examining gender-based violence can be informed by an understanding of how it develops among

adolescents in key social contexts such as schools. Many school programs designed to reduce this violence focus on the role of the bystander (e.g., Amar et al., 2012; Banyard et al., 2007; Cook-Craig et al., 2014; De La Rue et al., 2017; Midgett et al., 2015; Miller et al., 2012). This is because most school-based violence takes place in the presence of others (Polanin et al., 2012). Little is known about the effectiveness of these school-based interventions in the United Kingdom and how they impact on bystanders' attitudes, beliefs, motivations towards intervening, and intervention behavior when they witness gender-based violence. The current study is the first large-scale, multilevel evaluation of Mentors in Violence Prevention (Katz, 1995), a mentor-led program designed to challenge gender-based violence and violence more generally using a bystander approach, in the United Kingdom. It aimed to examine changes in bystander outcomes that were identified using a robust theoretical model of decision-making (Pagani et al., 2022a).

✉ Simon C. Hunter
Simon.Hunter@gcu.ac.uk

¹ University of Strathclyde, 16 Richmond Street, Glasgow G1 1XQ Scotland, UK

² Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA Scotland, UK

³ University of Western Australia, 35 Stirling Highway, Perth, WA 6009, Australia

⁴ Curtin University, Kent Street, Perth, WA 6102, Australia

Within the literature, gender-based violence has been studied in many forms ranging from non-intimate verbal and emotional violence, to dating and relationship abuse, to sexual violence (for recent reviews see De La Rue et al., 2017; Jouriles et al., 2018; Lester et al., 2017; Kettrey & Marx, 2019; Kovalenko et al., 2020). It is important to note that violence itself can manifest in different ways, involving different motivations from different perpetrators which are often complex and difficult to measure. For the purposes of the present study, the term gender-based violence reflects verbal and emotional violence which can involve direct and indirect forms such as name-calling, making fun, spreading rumors, ‘putting down’, and both physical and sexual violence, including pushing, hitting, fighting, sexting, touching, and coercive intercourse (Debnam & Mauer, 2021; Katz et al., 2011; Miller et al., 2012). These forms of violence can occur between boys and girls, both within and outside of romantic relationships. The term gender-based violence also incorporates domestic violence with regards to those in a romantic relationship as the sample explored in this thesis was under the age of 16 (Home Office, 2013).

Gender-based Violence and a Bystander Approach

Gender-based violence can reflect a motivation to assert power, whether that be within romantic relationships (Debnam & Mauer, 2021), or within peer groups (Skipper & Fox, 2021). This motivation for power can stem from stereotypical attitudes and normative beliefs pertaining to gender identity regarding females as subordinate in society and “lesser” than males (Hindes & Fileborn, 2020; Koss et al., 1994). Schools provide an ideal context in which to examine interactions and relationships between boys and girls because they are the dominant social setting for adolescents. Furthermore, adolescence is an important period in which to examine these relationships as it is a time when gender roles are explored and interpersonal relationships are negotiated (Katz, 2018; Skipper & Fox, 2021).

One way to examine stereotypical attitudes and normative beliefs in gender-based violence contexts is to examine how they manifest and develop in the attitudes, beliefs, and behaviors of those who are present when gender-based violence takes place. Indeed, research has identified that bystanders are present in 85–88% of violent situations in schools (Hawkins et al., 2001). How bystanders react is therefore key to understanding how situations of violence can be challenged since they have the power to facilitate (encourage the perpetrator) or inhibit (help the victim) a violent situation.

There are numerous school-based programs designed to promote bystander intervention in gender-based violence contexts (for meta-analyses and systematic reviews see: DeGue et al., 2014; Jouriles et al., 2018; Lundgren & Amin,

2015; Katz & Moore, 2013; Kettrey & Marx, 2019; Kovalenko et al., 2020; Stanley et al., 2015; Storer et al., 2016). Legislative, policy, and strategic advances have encouraged the implementation of these programs within school contexts. Indeed, the 2013 Campus Sexual Violence Elimination Act (Campus SaVE Act: 2013) made gender-based violence programs in the United States (US) of America mandatory. Most programs have therefore been pioneered and evaluated in the US, with fewer programs and evaluations elsewhere, such as the United Kingdom. Consequently, less is known about the effectiveness of gender-based violence programs in countries outside of the US (Kovalenko et al., 2020).

While the United Kingdom does not have legislation equivalent to the Campus SaVE Act, evaluations into the effectiveness of programs targeting gender-based violence do exist (e.g., Bell & Stanley, 2006; Fox et al., 2020; Hester & Westmarland, 2005; Maxwell et al., 2010; Scottish Executive, 2002; Stanley et al., 2011; Williams & Neville, 2013). However, evidence on the effectiveness of gender-based violence programs that focus on bystanders in secondary schools is scarce within the United Kingdom, with only one peer-reviewed qualitative evaluation existing (Williams & Neville, 2017). Other evaluations exist, but are either not peer-reviewed (e.g., Fox et al., 2020; Fox & Vickers, 2017; Mentors in Violence Prevention Progress Report, 2019; Williams & Neville, 2013) or do not specifically focus on bystander outcomes concerning gender-based violence (Hunter et al., 2021). Methodologically, there is also the need for larger evaluations that include both pre and post intervention testing, control schools, and designs that address the potential for school level differences. These issues are directly addressed in the current study.

Mentors in Violence Prevention

Mentors in Violence Prevention (Katz, 1995) was originally pioneered in universities in the US. It aims to empower bystanders to challenge negative gendered attitudes, beliefs, and behaviors. The original program was targeted towards male university sports athletes. The remit of Mentors in Violence Prevention has since evolved, and its implementation has expanded to settings such as high schools and involves training both boys and girls as role models. Mentors in Violence Prevention is also now often utilized to tackle violence beyond the original focus on gender-based violence (Katz, 2018). In Scotland, older students in schools (mentors) act as potential role-models for younger students (mentees), and lead younger students through a series of Mentors in Violence Prevention lessons. Mentors in Violence Prevention implements a “train the trainer” approach whereby school staff attend a two-day training course that is

run by development officers. Once they have received this initial training, school staff assume the position of Mentors in Violence Prevention leads and go on to train mentors (schools students) to implement lessons, utilising the same training that they themselves received (Mentors in Violence Prevention Progress Report, 2016). For each lesson, there is a lesson plan with a clear structure for mentors to follow. Lessons are usually incorporated into a normal school lesson with mixed gender classes (the size and structure depending on that of the school in question), replacing subjects such as Social Education. Lessons involve leading younger students through a hypothetical scenario which depicts a situation of gender-based violence or violence more generally. Situations include verbal, emotional, physical, and/or sexual violence. Following the presentation of the scenario, mentors engage the younger students in a “train of thought” about the scenario, allowing them to consider their beliefs, attitudes, and thought patterns. Finally, seven possible bystander reactions are presented, and they are invited to discuss the positive and negative consequences of each one. Mentors in Violence Prevention has five core values which mentors are encouraged to cover in each of their lessons: violence through a gendered lens, using a bystander approach, developing leadership, recognising the scope of violence, and challenging victim blaming.

Mentors in Violence Prevention has been implemented in the United Kingdom since 2012, first in Scotland and then in England in 2015. Both university and high school evaluations have been conducted in the US (Beardall, 2007; Cissner, 2009; Eriksen, 2015; Heisterkamp et al., 2011; Katz et al., 2011; Ward, 2001) and UK (Fox et al., 2020; Fox & Vickers, 2017; Hunter et al., 2021; Mentors in Violence Prevention Progress Report, 2019; Williams & Neville, 2013, 2017) respectively. Positive effects have been reported in relation to attitudes towards intervening (Beardall, 2007; Cissner, 2009; Eriksen, 2015; Fox et al., 2020; Heisterkamp et al., 2011; Katz et al., 2011; Mentors in Violence Prevention Progress Report, 2019; Ward, 2001; Williams & Neville, 2013, 2017), self-efficacy towards intervening (Beardall, 2007; Cissner, 2009; Eriksen, 2015; Fox & Vickers, 2017; Ward, 2001; Williams & Neville, 2017), and intentions to intervene (Cissner, 2009; Eriksen, 2015; Heisterkamp et al., 2011; Katz et al., 2011; Mentors in Violence Prevention Progress Report, 2019; Ward, 2001; Williams & Neville, 2017). Two studies also noted overall decreases in school rates of gender-based violence after the intervention (Cissner, 2009; Heisterkamp et al., 2011). However, not all studies have included control schools (Beardall, 2007; Eriksen, 2015; Fox et al., 2020; Fox & Vickers, 2017; Mentors in Violence Prevention Progress Report, 2019; Williams & Neville 2017), or pre and post testing methods (Beardall, 2007; Fox & Vickers, 2017;

Hunter et al., 2021; Katz et al., 2011; Mentors in Violence Prevention Progress Report, 2019; Williams & Neville, 2017), which are important in order to draw reliable comparisons (between similar schools or the same school at different time points) that provide a more robust and accurate insight into the effectiveness of an intervention (e.g., Kovalenko et al., 2020). Many studies also included small numbers (two or three) of intervention schools (Beardall, 2007; Cissner, 2009; Eriksen, 2015; Heisterkamp et al., 2011; Katz et al., 2011; Williams & Neville, 2013, 2017). Some of these studies that did not use control or pre and post testing methods, and included a small number of schools, found strong, positive Mentors in Violence Prevention effects or perceived effects (Beardall, 2007; Katz et al., 2011; Williams & Neville, 2013), whereas others found more mixed or even null effects (Eriksen, 2015; Fox et al., 2020; Hunter et al., 2021; Ward, 2001; Williams & Neville, 2013). Some of these latter studies also examined larger numbers of schools (Fox et al., 2020; Hunter et al., 2021; Ward, 2001). Nonetheless, other studies that have used control schools, did find positive intervention effects (Cissner, 2009; Heisterkamp et al., 2011), however, these studies included a small number of schools: two and four respectively.

Studies have examined the effectiveness of bystander intervention programs by focussing on the individual bystander and therefore changes in individual social cognitive factors that are influential when bystanders are making the decision over whether or not to intervene (e.g., Debnam & Mauer, 2021; Hoxmeier et al., 2018; Sjögren et al., 2021; Thornberg & Wänström, 2018). These factors are components of theoretical models of decision-making, such as the Theory of Planned Behavior (Ajzen, 1988, 1991) and the Prototype Willingness Model (Gibbons & Gerrard, 1995, 1997). Unfortunately, many of these studies have only examined some of the factors that comprise these theoretical models, leading to a partial picture of bystander decision-making. There is therefore a need for more studies to examine all factors that comprise theoretical models of bystander decision-making (Pagani et al., 2022a). Similarly, there is also a need for studies to assess changes in all these factors.

This Prototype Willingness Model has been tested to identify the predictors of bystander intervention behavior in gender-based violence contexts (Pagani et al., 2022a). Those constructs that were found to successfully predict bystander intervention were: willingness to intervene in less serious gender-based violence; positive attitudes (positive evaluations of intervening); negative attitudes (negative evaluations of intervening); subjective norms (beliefs about other bystanders intervening); self-efficacy (perceived ability to intervene); prototype perceptions (identification with the typical bystander who intervenes); moral

disengagement (beliefs about whether intervening is the right thing to do). Some of these factors (positive and negative attitudes towards intervening, prototype perceptions concerning self-comparison to the typical bystander who intervenes regularly, willingness to intervene) have been novelly tested in the context of bystander decision-making (Pagani et al., 2022a). However, these factors have been shown to predict decision-making in a range of other situations, including speeding, binge-drinking, smoking, and consuming a high-fat diet (Elliott et al., 2015; McCartan et al., 2018). Other factors (subjective norms concerning beliefs about other bystanders' intervention behaviors, self-efficacy concerning confidence in one's own ability to intervene, and moral disengagement by justifying gender-based violence) have also been extensively examined in the field of bystander decision-making (e.g., Rosval, 2013, Sjögren et al., 2021; Thornberg & Jungert, 2014). These factors have been examined within other theoretical models of bystander decision-making, such as the Theory of Planned Behavior (Ajzen, 1991).

The constructs that did not significantly predict bystander intervention behavior (Pagani et al. 2022a) were: intentions (likelihood to intervene); willingness to intervene in more serious gender-based violence; and perceived behavioral control (belief about whether intervening is under one's own control). However, willingness predicts behavior in a range of other situations (e.g., Elliott et al., 2015). Like subjective norms, perceived behavioral control is a construct of the Theory of Planned Behavior, however, it has been replaced or conflated with self-efficacy in the bystander literature (Rosval, 2013; Salmivalli, 2010; Sjögren et al., 2021). However, factor analytic studies have illustrated their independent effects in predicting intentions and behavior (Armitage & Conner, 1999a, 1999b). The predictive ability of intentions has been examined widely in different bystander situations (e.g., Leone & Parrott, 2019; McMahan et al., 2015; Rosval, 2013). Therefore, the inclusion of these factors would ensure a theoretically comprehensive range of social cognitions found to predict many other health-related behaviors (Elliott et al., 2015; Hoxmeier et al., 2018; McCartan et al., 2018; McCartan & Elliott, 2018; Rosval, 2013).

Current Study

The present study aimed to address the above stated limitations by examining changes in bystanders' attitudes, beliefs, motivations towards intervening, and intervention behavior after having received the Mentors in Violence Prevention intervention, and by using outcomes that comprise a robust theoretical model of decision-making. To achieve this, both control and pre and post testing methods

were used, with a 12-month follow-up to help provide a better insight into the mixed findings above in relation to the different methods used. In addition, effectiveness was tested using both measures of bystander intervention behavior and a comprehensive range of socio-cognitive constructs that comprise a robust theoretical model of decision-making. Specifically, this study used an augmented Prototype Willingness model. Those decision-making constructs above that were found to significantly predict bystander intervention behavior were deemed the most suitable outcome variables (due to their significant influence in bystander decision-making using the same sample as used in this study) for assessing the effectiveness of Mentors in Violence Prevention. They were therefore included in the confirmatory analyses in this study. In line with the pre-registration¹ for this study on the Open Science Framework (OSF: <https://doi.org/10.17605/OSF.IO/ZCJS6>), it was therefore hypothesized that: compared to those in non-intervention schools, students in Mentors in Violence Prevention schools would have significantly larger changes in these outcome variables and both positive intervention behavior (e.g., confronting the perpetrator of gender-based violence) and negative intervention behavior (e.g., doing nothing). The constructs above that did not significantly predict bystander intervention behavior were included in exploratory analyses in the current study.

Method

Participants

The final sample at both T1 and T2 was 1396 (50% attending Mentors in Violence Prevention (MVP) and 50% attending non-MVP schools) students attending 17 (9 MVP and 8 non-MVP) mainstream high schools in Scotland between Autumn 2018 and Spring 2020. Participants were aged 11 to 14 years old ($M = 12.25$, $SD = 0.84$). Overall, 689 (49.4%) reported their gender as female, 685 (49.1%) as male, 9 (0.6%) "prefer not to say", and 13 (0.9%) did not report. Ninety one percent of the sample identified as "White Scottish or White British" ($N = 1270$), 2.4% identified as "Asian, Asian Scottish/ British", 0.9% identified as "African", 1.9% identified as "Mixed or multiple ethnic group", 0.4% identified as "Caribbean or Black" and 2.9% identified as "Other ethnic group". Socioeconomic status, as assessed by the percentage of students in each school who

¹ Our preregistration on the Open Science Framework includes hypotheses, copies of all measures, relevant data, Mplus code, and other project materials such as participant consent and information sheets.

were registered for free school meals, ranged from 1.0% - 50.2% ($M = 20.6\%$, $SD = 14.0$).

In the preregistration (OSF: <https://doi.org/10.17605/OSF.IO/ZCJS6>), simulation analyses were conducted (Castelloe and O'Brien, 2001) to determine if the sample size was sufficient for the anticipated design of the study, that is, a multilevel, longitudinal design. The same sample recruited was used here as in another study (Pagani et al., 2022a), though the current study includes data from a second time point which occurred approximately 12 months after the first. T1 data therefore informed the simulation approach, which indicated that the sample sizes across anticipated outcomes were sufficient to detect meaningful effect sizes (small-medium effects $d = 0.25$ – 0.35 ; Cohen, 1992) across all the outcomes. However, the anticipated sample size ($N = 671$) for the more serious intervention behavior outcomes was considerably more than the observed sample size ($N = 288$).²

Design and Procedure

An observational, longitudinal design was used, where schools were pre-allocated to MVP and non-MVP groups. All anticipated outcome variables were measured at T1 (between Autumn 2018 and Spring 2019), and then subsequently again at T2, approximately one year later. Covariates (gender, age, ethnicity, empathy) were measured at T1 only.

A total of 2079 young people in S1-S3 attending 19 (14 MVP, 5 non-MVP) mainstream secondary schools in Scotland consented to take part at T1. However, between T1 and T2, five of the 'MVP' schools were unable to implement MVP due to staffing and structural issues: therefore, three subsequently became non-MVP schools and two withdrew entirely due to school closures related to the COVID-19 pandemic. In addition, and as a result of the pandemic, two more non-MVP schools had completed partial data collection before they were forced to close. Consequently, one school reported on intervention behavior only at T2 and the other school reported on all other variables only at T2. Both were included in the analyses. As described above, the final sample for T1 and T2 was 1396 students, with 50% (698) attending non-MVP schools and 50% attending MVP schools. However, participation within schools varied considerably due to whether positive or negative consent was sought as well as the availability of the target age groups at the time of fieldwork. Numbers of participants varied from 5 to 198 between schools.

² Single level linear regressions were also conducted on these outcomes, where the sample size was adequate to detect an effect. However, as was the case for the multilevel regressions (see results section), the linear regressions revealed that MVP had no significant effects on change scores for the more serious intervention outcomes (Negative intervention_{MoreSerious}: $\beta = -0.02$, $p = 0.847$; Positive intervention_{MoreSerious}: $\beta = -0.19$, $p = 0.076$).

Ethical approval for this study was granted by the lead author's institution. Scottish Local Education Authorities (LEAs) were contacted for permission to approach schools. Permission was obtained from 10 of 13 (77%) LEAs. All schools within the ten authorities were invited to take part. Of the schools contacted, 42 (26%) expressed interest in taking part and 19 did so. Time, resource limitations, and eligibility prevented some schools from participating. The research team aimed to recruit a balanced number of MVP and non-MVP schools. However, the fact that more MVP schools (14) opted into the study than non-MVP schools (5) potentially reflected the commitment of these schools to the MVP program or indeed their motivation to understand the effectiveness of a program that they had committed so much time and resource to. Schools could not be randomly assigned to these groups as the implementation of MVP is led by the Scottish Government and thus randomization was not within the control of the research team.

Information letters and consent forms were distributed to all parents whose children were eligible. At the preference of the LEA, either negative (80% of LEAs) or positive (20%) consent was sought. Parents were given at least one week to return consent. Each young person also consented before participation. Participants completed anonymous questionnaires within a classroom or assembly hall context. At both T1 and T2, data collection occurred in two phases: all measures (see section below) except intervention behavior were completed in the first phase and reported intervention behavior was assessed in a second phase approximately one month later. The questionnaire completed in phase 1 took one school period (45–55 min) to complete, and 5–10 min to complete in phase 2. Teachers and members of the research team were present during data collection. Participants were subsequently debriefed. Students who were not participating were usually given another classroom task to complete or attended a non-participating classroom that was undertaking their work as normal.

Measures

Scale items were drawn from the literature (e.g., Miller et al., 2012) for each measure. Eight gender-based violence examples (Miller et al., 2012; see Appendix 1 for a description of the eight examples, divided into their verbal/emotional and physical/sexual counterparts) were integrated within each measure. These included examples of emotional, verbal, physical, and/or sexual violence. All eight gender-based violence examples were not used to assess every construct to reduce burden, but examples were included so that a balanced range of verbal/emotional and physical/sexual violence items assessed each. For most measures (positive and negative attitudes towards intervening, self-efficacy, perceived behavioral control over intervening, subjective norms concerning

beliefs about other bystanders' intervention behaviors, prototype perceptions concerning self-comparison to the typical bystander who intervenes regularly, willingness to intervene), questions were split into parts "a" and "b", where part "a" reflected an example of verbal/emotional violence and part "b" reflected an example of physical/sexual violence. For measures that were structured differently (intentions to intervene, self-reported behavior, moral disengagement by justifying gender-based violence), a balanced mix of verbal/emotional (4 items for intentions and self-reported behavior, and 3 items for moral disengagement) and physical/sexual (4 items for intentions and self-reported behavior, and 3 for moral disengagement) examples were incorporated into the questions. Participants were thus exposed equally to the eight examples. The only measure that did not take this approach was empathy and this was because it measured a personality trait. All outcome variables were measured using unipolar scales in line with standard practice (e.g., Fishbein & Ajzen, 2010). The factorial structures of the measures follow that in another study (Pagani et al., 2022a) and so, factor scores were created for each measure by summing each item's raw scores multiplied by that item's factor weight.

Outcome variables for confirmatory analyses

Attitudes towards intervening

Both positive and negative attitudes were measured separately (Elliott et al., 2015). Six items were used to measure positive attitudes towards intervening (e.g., How positive would it be if you did something about it if you saw [violence example]). The participants were asked to respond to each item using a 9-point unipolar scale (e.g., 1 = *not at all positive* to 9 = *extremely positive*). Similarly, six items were used to measure negative attitudes towards intervening (e.g., How negative would it be if you did something about it if you saw [violence example]). The participants were asked to respond to each item using a 9-point unipolar scale (e.g., 1 = *not at all negative* to 9 = *extremely negative*). For both the positive and negative scales, each of the six items were divided into "a" and "b" components, essentially making each of the scales consist of 12 indices. Factor scores were created where a higher score indicated higher positive or higher negative attitudes. Internal reliability was high at T1 (positive scale $\alpha = 0.93$; negative scale $\alpha = 0.93$) and T2 (positive scale $\alpha = 0.93$; negative scale $\alpha = 0.92$).

Subjective norms concerning beliefs about other bystanders intervening

A three-item scale was adapted (Wilson et al., 2016). Participants rated responses to questions like, "Of the students

you know, how many do you think will do something about it over the next month when they see... (*violence example*)", from 1 = *none of them* to 9 = *all of them*. For this scale, each of the three items were divided into "a" and "b" components, essentially making the scale consist of six items. Factor scores were created, and a higher score indicated higher levels of perceived social pressure to intervene. Internal reliability was high at T1 ($\alpha = 0.82$) and T2 ($\alpha = 0.79$).

Self-efficacy concerning confidence in one's own ability to intervene

A three-item scale was adapted (Wilson et al., 2016). Participants rated responses to questions like, "Over the next month, I have the ability to do something about it when I see... (*violence example*)", from 1 = *not at all confident* to 9 = *very confident*. For this scale, each of the three items were divided into "a" and "b" components, essentially making the scale consist of six items. Factor scores were generated where a higher score indicated higher self-efficacy to intervene. Internal reliability was high at T1 ($\alpha = 0.75$) and T2 ($\alpha = 0.73$).

Prototype perceptions concerning self-comparison to the typical bystander who intervenes regularly

A four-item scale was adapted (Elliott et al., 2017). Participants rated responses to questions like "Do you resemble the type of person your age that regularly does something about it when they see... (*violence example*)", from 1 = *definitely no* to 9 = *definitely yes*. For this scale, each of the four items were divided into "a" and "b" components, essentially making the scale consist of eight items. Factor scores were created, and higher scores indicated that participants regarded themselves as being more like the perceived prototypical student who intervenes regularly. Internal reliability was high at T1 ($\alpha = 0.91$) and T2 ($\alpha = 0.90$).

Moral disengagement concerning justification of gender-based violence

A six-item scale was adapted (Thornberg & Jungert, 2014). Participants rated responses to questions like, "It's okay for... (*violence example*)", from 1 = *strongly agree* to 7 = *strongly disagree*. As described at the start of this section, three of the items included examples of verbal/emotional violence and the other three included examples of physical/sexual violence. Factor scores were created, and items were reverse scored so that a higher score indicated higher moral disengagement. Internal reliability was high at T1 ($\alpha = 0.91$) and T2 ($\alpha = 0.90$).

Table 1 Proportions of participants who had witnessed between 0 and 4 situations of more and less serious gender-based violence at T1 and T2 in MVP, non-MVP, and overall groups

		More serious gender-based violence			Less serious gender-based violence			Count
		No opportunity	Saw 1–2 situations	Saw 3–4 situations	No opportunity	Saw 1–2 situations	Saw 3–4 situations	
MVP	T1	60.28%	33.51%	6.21%	46.46%	41.11%	12.43%	557
	T2	57.86%	35.23%	6.91%	34.02%	49.91%	16.07%	
NON-MVP	T1	54.58%	38.06%	7.36%	26.57%	44.70%	28.73%	579
	T2	53.68%	36.80%	9.52%	20.83%	46.32%	32.85%	
OVERALL	T1	57.50%	35.74%	6.76%	36.71%	42.87%	20.42%	1136
	T2	55.81%	36.00%	8.19%	27.55%	48.15%	24.30%	

These proportions only reflect whether or not bystanders had the opportunity to intervene in each of the eight situations of gender-based violence and not the rates of opportunities they had within each of the situations. Exact numbers are not reported on as some of the data did not contain numerical estimates, but instead descriptive text that reflected the amount of times they had witnessed each situation, for example, “lots” or “too many to count”.

Behavioral willingness to intervene

A three-item scale was adapted (Elliott et al., 2017). Participants rated responses to questions like, “Suppose you saw...(*violence example*)...over the next month and no-one else there was doing anything about it/ none of your friends were intervening/ no-one else was around. How willing would you be to do something?” (1 = *not at all willing* to 9 = *extremely willing*). For this scale, each of the three items were divided into “a” and “b” components, essentially making the scale consist of six items. The factor analytic approach revealed that this measure constituted “less serious” and “more serious” gender-based violence components, where “a” items loaded onto one factor and “b” items loaded onto another. These were therefore included as separate measures. Internal reliability was high for T1 (willingness_{MoreSerious}: $\alpha = 0.82$; willingness_{LessSerious}: $\alpha = 0.86$), and T2 (willingness_{MoreSerious}: $\alpha = 0.88$; willingness_{LessSerious}: $\alpha = 0.89$).

Self-reported intervention behavior

Mentees reported if they had witnessed eight gender-based violence situations in the previous month (Miller et al., 2012). For each situation witnessed, participants then reported how they had intervened by ticking box(es) reflecting two negative (e.g., “I didn’t do/say anything”), and four positive (e.g., “I told the person in public that acting like that was not ok”) responses (Miller et al., 2012). As described at the start of this section, four of the items included examples of verbal/emotional violence and the other four included examples of physical/sexual violence. This measure also contained more serious and less serious components of gender-based violence (Pagani et al., 2022a), resulting in four measures of intervention behavior: the proportion of times the mentees intervened positively in less serious situations (positive intervention_{LessSerious}), that which they intervened positively in more serious situations (positive

intervention_{MoreSerious}), that which they intervened negatively in less serious situations (negative intervention_{LessSerious}) and that which they intervened negatively in more serious situations (negative intervention_{MoreSerious}).

There were 1136 mentees who responded to the self-reported bystander behavior questionnaire at both T1 and T2. This is slightly smaller than the reported number of participants ($n = 1396$) and could likely reflect student absences as well as refusals to complete the questionnaires on the date that this fieldwork was completed, which was around one month after the longer questionnaire, as described in the earlier design and procedure. Table 1 shows the proportions of youth who reported witnessing the eight possible gender-based violence situations examined in this report. These are split by both time points and by MVP vs. non-MVP schools.

Covariates

Gender

Gender was reported as “a boy”, “a girl”, or “prefer not to say”. Those who responded “prefer not to say” or with missing responses were omitted due to small numbers (1.5% of total responses). Gender was coded as 0 = *boy* and 1 = *girl* for analyses.

Age

Ages ranged from 11–15. No treatment of this variable occurred for analyses.

Ethnicity

Participants reported their ethnicity as “White Scottish or White British”, “Asian, Asian Scottish/ British”, “African”, “Mixed or multiple ethnic group”, “Caribbean or Black”,

and “Other ethnic group”. Due to small numbers in each of these categories of response aside from “White Scottish or White British” (see participants section), all other ethnic groups were collapsed into one group “other ethnic group”. Ethnicity was coded as 0 = *White Scottish or White British*, and 1 = *other ethnic group* for analyses.

Empathy

A six-item scale was used (Caravita et al., 2009), where participants rated responses to questions like, “Seeing a friend crying makes me feel as if I am crying”, from 1 = *never true* to 4 = *always true*. A mean score was created where a higher score indicated higher affective empathy. Internal reliability was satisfactory ($\alpha = 0.77$).

Outcome variables for exploratory analyses

Perceived behavioral control concerning perceptions of one’s control over intervening

A two-item scale was adapted (Wilson et al., 2016). Participants rated responses to questions like, “Over the next month, how much personal control do you feel you have over doing something about it when you see... (*violence example*)”, from 1 = *no control at all* to 9 = *complete control*. For this scale, each of the two items were divided into “a” and “b” components, essentially making the scale consist of four items. Factor scores were created, and a higher score indicated higher perceived control over intervening. Internal reliability was high at T1 ($\alpha = 0.88$) and T2 ($\alpha = 0.90$).

Intentions to intervene

An eight-item scale was adapted (Miller et al., 2012). Participants rated responses to questions like “How likely are you to do something about it over the next month if a male peer / friend is... (*violence example*)”, from 1 = *very unlikely* to 5 = *very likely*. As described at the start of this section, four of the items included examples of verbal/emotional violence and the other four included examples of physical/sexual violence. Factor scores were generated where a higher score indicated higher intentions to intervene. Internal reliability for this scale was high at T1 ($\alpha = 0.95$) and T2 ($\alpha = 0.93$).

Analytic Plan

A series of confirmatory multilevel linear regressions were planned for the anticipated outcomes. A multiple indicator factor-analytic approach was preregistered to be used to compute latent change scores representing outcome

variables. However, the number of parameter estimates (116) outnumbered the number of clusters (16 schools) in the study, and so a more parsimonious model was required. An observed change score was thus computed (Castro-Schilo & Grimm, 2018) for each outcome variable by subtracting factor scores at T1 from those at T2. The predictor was MVP (0 = *non-MVP*, 1 = *MVP*), and the covariates were gender, age, ethnicity, and affective empathy. The multilevel approach allowed within school effects to be estimated using intraclass correlations. Three exploratory multilevel linear regressions were also planned. Due to the large number of analyses, alpha for exploratory analyses was adjusted to $p < 0.01$.

Results

Treatment of Missing Data

Participants with any partial missing data across the outcome variables were deleted, therefore, final sample sizes within each analysis varied. With the exception of the self-reported intervention behavior outcome variables, data were missing completely at random, and sample sizes varied from 1075 (23.0% missingness) to 1223 participants (12.4% missingness). For the self-reported intervention behavior outcome variables, missing data was determined by whether or not participants had the opportunity to intervene in more serious or less serious gender-based violence in the preceding month (see measures section). This was true for both T1 and T2, so some participants who had the opportunity to intervene at T1 may not have had the opportunity at T2 and vice-versa. For the more serious intervention behavior outcome variables, the sample size was 288 (79.4% missingness) who had the opportunity to intervene at both T1 and T2. For the less serious intervention behavior outcome variables, the sample size was 618 (55.5% missingness) who had the opportunity to intervene at both T1 and T2. For the covariates, a Full Information Maximum Likelihood (FIML) approach was implemented in MPlus to address any missing data.

Descriptive Statistics

Table 2 shows the mean change scores and intraclass correlation (ICC) coefficients for all outcome variables, representing those for MVP and non-MVP schools separately as well as those for MVP and non-MVP schools combined.

The differences between T1 and T2 scores across all outcomes were not statistically significant for non-MVP schools. For MVP schools, negative attitudes towards intervening, $t(df = 565) = 2.91, p = 0.002, d = -0.12$, and moral disengagement by justifying gender-based violence, t

Table 2 Mean change score and ICCs for confirmatory outcomes across schools

	MVP Schools Change Scores		Non-MVP Schools Change Scores		Overall Change Scores	
	Mean	ICC	Mean	ICC	Mean	ICC
positive attitudes ⁺	0.05	0.00	-0.04	0.01	0.00	0.01
negative attitudes ⁻	-0.16	NC	-0.07	0.00	-0.11	0.01
subjective norms ⁺	-0.02	0.00	-0.02	0.02	-0.02	0.01
self-efficacy ⁺	-0.02	0.00	-0.01	NC	-0.01	0.00
prototype perceptions ⁺	-0.03	0.02	-0.00	NC	-0.01	0.02
moral disengagement ⁻	-0.11	NC	-0.06	NC	-0.08	0.00
willingness _{LessSerious} ⁺	-0.06	0.00	-0.04	0.03	-0.05	0.01
positive intervention _{LessSerious} ⁺	0.02	NC	0.03	0.06	0.02	0.02
positive intervention _{MoreSerious} ⁺	-0.09	NC	0.05	0.09	-0.02	0.05
negative intervention _{LessSerious} ⁻	-0.04	0.01	-0.04	NC	-0.04	0.00
negative intervention _{MoreSerious} ⁻	0.01	0.08	-0.02	0.04	-0.01	0.05

NC Non computable (can occur when ICCs are extremely small), ICC intraclass correlation coefficient, willingness_{LessSerious} = willingness to intervene in less serious gender-based violence; positive intervention_{LessSerious} = positive intervention in less serious gender-based violence; positive intervention_{MoreSerious} = positive intervention in more serious gender-based violence; negative intervention_{LessSerious} = negative intervention in less serious gender-based violence; negative intervention_{MoreSerious} = negative intervention in more serious gender-based violence. ⁺ = if the mean change score is positive then this indicates an improvement from T1 to T2. ⁻ = if the mean score is negative then this indicates an improvement from T1 to T2

(df = 548) = 2.55, $p = 0.006$, $d = -0.11$, improved from T1 to T2. However, positive intervention_{MoreSerious} (positive intervention in more serious gender-based violence) deteriorated from T1 to T2, $t(df = 135) = 1.85$, $p = 0.033$, $d = -0.16$.

ICC scores ranged from 0.00 to 0.09 (Median = 0.01) in non-MVP schools and 0.00–0.08 (Median = 0.08) in MVP schools, indicating that some outcomes (e.g., prototype perceptions, negative_{MoreSerious}) had higher similarity among those within the same schools than others (e.g., negative attitudes towards intervening, self-efficacy concerning confidence in one's own ability to intervene), and that non-MVP schools generally had more outcomes showing higher similarity between those in the same school than MVP schools.

Confirmatory Regression Analyses

Tables 3 and 4 present the results of the confirmatory regression analyses testing the effects of MVP on the change scores for the outcome variables.

As can be seen in Tables 3 and 4, the hypotheses were not supported: MVP did not significantly predict any change score. Regarding covariates, gender had no significant effect. Ethnicity significantly predicted change scores in positive attitudes, where those who are not “White Scottish/White British” had larger improvements in positive attitudes towards intervening in gender-based violence. Age significantly predicted change scores in subjective norms (beliefs about whether others would intervene), negative intervention_{LessSerious} (negative intervention in less serious violence), and positive intervention_{LessSerious} (positive intervention in less serious violence). Specifically, as age increased the change scores in subjective norms concerning beliefs about other bystander's intervention behaviors deteriorated, and the change scores in positive and negative intervention_{LessSerious} improved. Empathy also predicted self-efficacy, so as empathy increased, self-efficacy concerning confidence in one's own ability to intervene improved.

Exploratory Regression Analyses

Perceived behavioral control over intervening increased significantly from T1 to T2 in both MVP, $t(df = 553) = -3.63$, $p < 0.001$, $d = 0.15$, and non-MVP, $t(df = 596) = -3.46$, $p < 0.001$, $d = 0.14$, schools. ICC scores were 0.01 for MVP schools and 0.00 for non-MVP schools. MVP exposure did not have an effect on this variable's change score ($B = 0.03$, $p = 0.737$), suggesting that those in MVP schools did not change significantly in their levels of perceived behavioral control over intervening compared to those in non-MVP schools. There was also no effect for the covariates.

Behavioral intentions to intervene increased significantly from T1 to T2 in both MVP, $t(df = 547) = -4.56$, $p < 0.001$, $d = 0.20$, and non-MVP, $t(df = 611) = -2.89$, $p = 0.002$, $d = 0.12$, schools. It was not possible to compute ICC scores (potentially because they were extremely small) for MVP schools though this was 0.01 for non-MVP schools. MVP exposure did not have an effect on this variable's change score ($B = 0.10$, $p = 0.135$), however gender ($B = -0.21$, $p < 0.001$) and empathy ($B = 0.11$, $p = 0.007$) did. This suggests that those in MVP schools did not change significantly in intentions compared to those in non-MVP schools, that intentions to intervene improved more for boys and that, as empathy increased, intentions to intervene improved more.

Willingness_{MoreSerious} to intervene did not change significantly from T1 to T2 in MVP (Mean_{Change} = -0.01, $p = 0.424$), but increased in non-MVP schools, $t(df = 572) = -1.79$, $p = 0.033$, $d = 0.10$, schools. ICC scores were 0.03 in MVP schools and 0.02 in non-MVP schools. MVP exposure did not have an effect on the change score for this variable

Table 3 Unstandardized estimated effects on change scores for outcome variables, including within and between effects

	positive attitudes		negative attitudes		self-efficacy		subjective norms		prototype perceptions		moral disengagement	
	B	p-value	B	p-value	B	p-value	B	p-value	B	p-value	B	p-value
<i>Within level</i>												
Gender	0.06	0.498	-0.06	0.175	-0.01	0.850	-0.08	0.142	0.00	0.994	-0.02	0.778
Ethnicity	0.24	0.020	0.05	0.694	0.02	0.643	0.03	0.564	-0.10	0.676	-0.01	0.944
Empathy	-0.07	0.455	0.02	0.812	0.07	0.025	-0.04	0.376	-0.02	0.776	-0.01	0.919
Age	-0.01	0.793	-0.07	0.140	0.02	0.356	-0.03	0.041	0.08	0.268	0.03	0.607
<i>Between level</i>												
MVP intervention	0.10	0.112	-0.03	0.556	-0.00	0.997	-0.01	0.969	0.03	0.827	-0.10	0.170

Table 4 Unstandardized estimated effects on change scores for outcome variables, including within and between effects

	Willingness _{LessSerious}		Positive intervention _{MoreSerious}		Positive intervention _{LessSerious}		Negative intervention _{MoreSerious}		Negative intervention _{LessSerious}	
	B	p-value	B	p-value	B	p-value	B	p-value	B	p-value
<i>Within level</i>										
Gender	0.03	0.646	0.05	0.627	0.05	0.095	0.03	0.668	0.03	0.429
Ethnicity	-0.11	0.074	0.14	0.349	0.00	0.958	0.03	0.814	-0.06	0.435
Empathy	0.01	0.838	0.00	0.967	0.04	0.327	-0.02	0.822	-0.07	0.328
Age	-0.02	0.711	-0.03	0.783	0.07	0.012	-0.07	0.256	-0.06	0.001
<i>Between level</i>										
MVP intervention	-0.07	0.646	-0.03	0.796	0.02	0.674	0.03	0.645	-0.02	0.625

willingness_{LessSerious} = willingness to intervene in less serious gender-based violence; positive intervention_{LessSerious} = positive intervention in less serious gender-based violence; positive intervention_{MoreSerious} = positive intervention in more serious gender-based violence; negative intervention_{LessSerious} = negative intervention in less serious gender-based violence; negative intervention_{MoreSerious} = negative intervention in more serious gender-based violence

(B = 0.03, p = 0.737), suggesting that those in MVP schools did not change significantly in their levels of willingness_{MoreSerious} compared to students in non-MVP schools. There were also no effects for the covariates.

Discussion

There are few peer-reviewed studies that contain pre- and post-intervention and control methods to examine the effectiveness of programmes aimed at tackling gender-based violence by using a bystander intervention approach in the United Kingdom. Furthermore, existing studies, including those that are conducted outside of the United Kingdom, tend to provide a partial examination of bystander decision-making factors, rather than a full set of factors that comprise a robust theoretical model of decision-making. This study aimed to address these issues by providing insight into changes in bystander attitudes, beliefs, motivations, and behaviors when violence is witnessed among Scottish high school students. These outcomes were identified using a robust theoretical model of decision-making (Pagani et al., 2022a) which informed the Mentors in Violence Prevention intervention program.

This intervention takes a mentor-led approach to target bystanders’ attitudes, beliefs, motivations towards intervening, and intervention behavior in different situations of gender-based violence. This study was the first large-scale multilevel evaluation in the United Kingdom. Intervention exposure did not significantly impact on anticipated outcomes, including positive and negative attitudes towards intervening, self-efficacy concerning confidence in one’s own ability to intervene, subjective norms concerning beliefs in other bystanders’ intervention behaviors, prototype perceptions concerning self-comparison to the typical bystander who intervenes regularly, positive intervention behavior, or negative intervention behavior. Neither were there between school variations in the effect of the intervention. Exploratory analyses revealed that Mentors in Violence Prevention did not have an effect on perceived behavioral control over one’s own intervention behavior, intentions to intervene, or willingness to intervene in more serious gender-based violence. The results do not therefore provide any evidence that Mentors in Violence Prevention was effective at modifying the anticipated outcomes pertaining to bystander decision-making and intervention in different situations of gender-based violence.

These findings are in line with other research finding no evidence for the effectiveness of Mentors in Violence Prevention (Hunter et al., 2021). Nevertheless, it should be noted that the majority of relevant evaluation research has found some positive effects for the program. Specifically, and in direct contrast to this study, significant differences in self-efficacy concerning confidence in one's own ability to intervene (Beardall, 2007; Cissner, 2009; Eriksen, 2015; Fox & Vickers, 2017; Ward, 2001; Williams & Neville, 2017), and intentions to intervene (Cissner, 2009; Eriksen, 2015; Heisterkamp et al., 2011; Katz et al., 2011; MVP Progress Report, 2019; Ward, 2001; Williams & Neville, 2017) have been reported.

Methodological differences between the current study and Mentors in Violence Prevention evaluations reporting positive effects may partially explain the discrepancies between the results. This study included a large number of schools from multiple Local Education Authorities across Scotland. Many of the other evaluations examined the effectiveness of Mentors in Violence Prevention in only one (Beardall, 2007; Cissner, 2009; Eriksen, 2015), two (Katz et al., 2011), or three schools (Heisterkamp et al., 2011; Williams & Neville, 2013, 2017). Indeed, studies that evaluated intervention effectiveness across a larger number of schools tended to yield more mixed or null findings (Fox et al., 2020; Hunter et al., 2021; Ward, 2001). Thus, it is possible that studies involving only a very small number of schools may also have recruited very committed schools who are more effective at producing change as a result.

Cultural variations may also help to explain the differences in findings between the evaluations conducted in America and England and those conducted in Scotland. For example, the Mentors in Violence Prevention program was adapted specifically for a Scottish context (Mentors in Violence Prevention Progress Report, 2016) by referring to knife crime rather than gun crime, which may mean that implementation in Scotland differs from that in America. Although, adaptations to suit the cultural context should be beneficial (Durlak & DuPre, 2008). Indeed, Mentors in Violence Prevention leads in schools are encouraged to focus on violent situations in lessons that are relevant to the school's own culture. This could potentially mean that leads choose to focus on very specific situations of violence, that may not include gender-based violence specifically or may not fully cover all the examples of gender-based violence that they were asked about in this study. Furthermore, when developing the research questions for this study, the gender-based violence situations that were included aligned to those discussed within Mentors in Violence Prevention lessons specifically.

Another potential explanation to consider for the null findings is implementation variation (DeGue et al., 2014; Durlak & DuPre, 2008; Jouriles et al., 2018; Kovalenko

et al., 2020; Storer et al., 2016). However, published work utilising the same sample used in this article (Pagani et al., 2022b) examining the implementation of Mentors in Violence Prevention, found no dosage (level of exposure to the program), fidelity (extent to which core components such as gender-based violence and bystander intervention were covered during lessons), or adaptation (extent to which components of the program were adapted) effects. However, the change of Mentors in Violence Prevention's scope to also focus on more general violence (Katz, 2018) could potentially dilute the communication of gender-based violence, essentially making it a tick-box exercise (Pagani et al., 2022b), given that mentors are encouraged to cover this as a core value of the intervention. This could be another example of cultural differences across Scotland and America, where the evaluations conducted in America may have been conducted within schools which were fully supportive of the primary focus being on gender-based violence. Evidence has also emerged that some schools in England have not directly addressed gender-based violence at all when implementing Mentors in Violence Prevention (Fox et al., 2020).

Systematic reviews and meta-analyses often report mixed effects, with a large number of programs having no effects (e.g., DeGue et al., 2014; Jouriles et al., 2018; Lundgren & Amin, 2015; Katz & Moore, 2013; Kettrey & Marx, 2019; Kovalenko et al., 2020; Stanley et al., 2015; Storer et al., 2016). A common explanation put forward here is that many gender-based violence programs lack a strong theoretical framework (Kovalenko et al., 2020). This is not the case with Mentors in Violence Prevention, which has strong underpinnings in social norms theory (Perkins & Berkowitz, 1986) and social cognitive models of decision-making (Ajzen, 1988, 1991; Gibbons & Gerrad, 1995, 1997). In addition, using the same sample and measures as this study, many socio-cognitive predictors have been shown to be significantly associated with intervention behavior (Pagani et al., 2022a). However, the extent to which these theoretical factors are explicitly addressed during lessons is less clear (Pagani et al., 2022b). Furthermore, how these factors are being framed within the context of the key messages that Mentors in Violence Prevention intends to communicate is unknown.

Strengths of the current study include the use of a longitudinal and multilevel design, and a robust theoretical framework, to examine intervention effects. However, the study was slightly underpowered to be able to detect meaningful effect sizes in some of the outcomes (namely, those pertaining to intervention behavior) as determined in the preregistered simulation analyses (OSF: <https://doi.org/10.17605/OSF.IO/ZCJS6>). Nonetheless, the design was sufficiently powered for all other anticipated outcomes, and the researchers endeavored to

maintain consistency in the analyses used. Publication of null findings is an important step in building a comprehensive evidence base for this, or any, intervention. They inform future refinements in the application of interventions (Miller-Halegoua, 2017) and, since science cannot self-correct without them, they support progress more generally (Munafò & Neill, 2016). Second, although this study did include a multilevel component at the school level, it would have been useful to also include this at the classroom level to examine cross-classroom effects, where classroom culture could be captured. Indeed, other researchers have found larger effects at the classroom level than at the school level (Kärnä et al., 2011), suggesting that individuals belonging to the same classroom have more similar changes than those belonging to the same school. Unfortunately, the current study was neither statistically powered nor designed to include this further clustering step within the multilevel analysis. We were also unable to record the exact times at which implementation schools began and ended the intervention. This meant that the time between completing the intervention and undertaking the T2 assessments may have varied across schools, introducing undesirable variation in results. Furthermore, the fact that three of the Mentors in Violence Prevention schools became non-intervention schools due to their inability to implement the intervention, does raise questions about these schools' capacity and willingness to implement the program, and could indicate a potential redistribution bias in the findings relating to non-intervention schools.

Conclusion

Little is known about the effectiveness of bystander programmes targeting gender-based violence in the United Kingdom. Furthermore, studies tend to include a partial examination of bystander outcomes that do not fully comprise robust theoretical models of decision-making. The current study aimed to address these limitations by conducting a rigorous evaluation of the Mentors in Violence Prevention (Katz, 1995) program in Scottish high schools. Non-significant results revealed that Mentors in Violence Prevention was not found to influence changes after a 12-month follow-up in students' self-reported intervention behaviors nor across any of the other anticipated outcomes, including attitudes, beliefs, motivations towards intervening, and willingness to intervene in less serious situations of gender-based violence. Multilevel modeling found that schools did not significantly differ in their changes, suggesting no between school variations in Mentors in Violence Prevention effects. Furthermore, exploratory analyses revealed that Mentors in Violence

Prevention was not effective in positively influencing perceived behavioral control, intentions to intervene, and willingness to intervene in more serious situations of gender-based violence. This suggests that the intervention was not effective at modifying the anticipated outcomes in this study. The implications discussed indicate that decisions need to be made at stakeholder level as to whether Mentors in Violence Prevention should be framed according to its original intentions, that is, to tackle gender-based violence, and the likely implications of not doing so. The null results of the current study could further be attributed to a failure to adequately address the theoretical model underpinning Mentors in Violence Prevention in practice. These issues are reminiscent of wider issues in evaluating interventions, and should be addressed before concluding that an intervention is ineffective at bringing about its anticipated changes.

Authors' Contributions S.P. conceived of the study, led its design and coordination, led the analysis, and drafted the manuscript; S.C.H. participated in the study design and coordination, and in interpretation of the data; M.A.E. participated in the study design and coordination, and in interpretation of the data; D.L. participated in the study design and analytic strategy, and in interpretation of the data. All authors read and approved the final manuscript.

Funding This study was conducted as part of a PhD Studentship award to SP, and was funded equally by The University of Strathclyde and the Scottish Violence Reduction Unit.

Data Sharing Declaration The datasets generated and/or analyzed during the current study are available on the Open Science Framework (OSF) at <https://doi.org/10.17605/OSF.IO/ZCJS6>.

Compliance with Ethical Standards

Conflict of Interest The Scottish Government is responsible for funding implementation of the Mentors in Violence Prevention intervention, and S.P. is now employed by the Scottish Government as a Senior Research Officer. However, SP's role is separate to this and does not directly involve a commissioning or implementation role regarding Mentors in Violence Prevention. S.C.H., M.A.E., and D.L. report no conflict of interests.

Preregistration The ethics application, letters to Local Education Authorities and schools, Participant Information Sheets, consent forms, and the surveys used to support the findings of this study are openly available on the Open Science Framework (OSF) at <https://doi.org/10.17605/OSF.IO/ZCJS6>.

Ethical Approval This research was granted ethical approval at the lead author's institution. The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki.

Informed Consent Data were collected on an informed, parental consent and student assent basis. Parents were sent a letter containing a description of the study and of the procedures. Depending on the preference of the Local Education Authority (a gatekeeper to accessing schools in Scotland), parents who did not wish their child to participate

in the survey either (i) returned a signed letter, indicating their desire to decline their child's participation, or (ii) returned a signed letter, indicating whether they did or did not wish their child to participate.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1

Table 1 Eight gender-based violence examples used from Miller et al. (2012)

Type of violence	Gender-based violence example
Verbal/emotional	A male peer/friend making rude or disrespectful comments about a girl's body, clothes, or makeup.
	A male peer/friend doing unwelcome or uninvited things toward a girl (or group of girls), such as howling, whistling, or making sexual gestures.
	A male peer/friend spreading rumors about a girl's sexual reputation, like saying she's 'easy to get with'.
	A male peer/friend telling sexual jokes that disrespect women and girls.
Physical/sexual	A male peer/friend showing other people sexual messages or naked/sexual pictures of a girl on a mobile phone or the internet.
	A male peer/friend arguing with a girl where he's starting to swear at or threaten her.
	A male peer/friend shoving, grabbing, or otherwise physically hurting a girl.
	A male peer/friend taking sexual advantage of a girl (like touching, kissing, having sex with) who is drunk or high from drugs.

References

- Ajzen, I. (1988). *Attitudes, personality, and behavior*. Chicago: IL: Dorsey.
- Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-t](https://doi.org/10.1016/0749-5978(91)90020-t).
- Amar, A. F., Sutherland, M., & Kesler, E. (2012). Evaluation of a bystander education program. *Issues in Mental Health Nursing*, 33(12), 851–857. <https://doi.org/10.3109/01612840.2012.709915>.
- Armitage, C. J., & Conner, M. (1999a). Distinguishing perceptions of control from self-efficacy: Predicting consumption of a low-fat diet using the theory of planned behavior. *Journal of Applied Social Psychology*, 29(1), 72–90. <https://doi.org/10.1111/j.1559-1816.1999.tb01375.x>.
- Armitage, C. J., & Conner, M. (1999b). The theory of planned behavior: Assessment of predictive validity and "perceived control. *British Journal of Social Psychology*, 38(1), 35–54. <https://doi.org/10.1348/014466699164022>.
- Banyard, V. L., Moynihan, M. M., & Plante, E. G. (2007). Sexual violence prevention through bystander education: An experimental evaluation. *Journal of Community Psychology*, 35(4), 463–481. <https://doi.org/10.1002/jcop.20159>.
- Beardall, N. G. (2007). *A program evaluation research study on the implementation the mentors in violence prevention program in a public high school: A dissertation*. Graduate School of Arts and Social Sciences, Lesley University. http://digitalcommons.lesley.edu/expressive_dissertations/32.
- Bell, J., & Stanley, N. (2006). Learning about domestic violence: Young people's responses to a Healthy Relationships program. *Sex Education*, 6(3), 237–250. <https://doi.org/10.1080/14681810600836356>.
- Castro-Schilo, L., & Grimm, K. J. (2018). Using residualized change versus difference scores for longitudinal research. *Journal of Social and Personal Relationships*, 35(1), 32–58. <https://doi.org/10.1177/0265407517718387>.
- Campus SaVE act (2013) Violence against women reauthorization act of 2013. pub. L. 113–114, 4, 127 stat. 54.
- Caravita, S., DiBlasio, P., & Salmivalli, C. (2009). Unique and interactive effects of empathy and social status on involvement in bullying. *Social Development*, 18(1), 140–163. <https://doi.org/10.1111/j.1467-9507.2008.00465.x>.
- Castelloe, J. M., & O'Brien, R. G. (2001). Power and sample size determination for linear models. In *Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference* (pp. 240–26).
- Cissner, A. B. (2009). Evaluating the mentors in violence prevention program. *Center for Court Innovation*, 451–468.
- Cook-Craig, P. G., Coker, A. L., Clear, E. R., Garcia, L. S., Bush, H. M., Brancato, C. J., Williams, C. M., & Fisher, B. S. (2014). Challenge and opportunity in evaluating a diffusion-based active bystander prevention program: Green Dot in high schools. *Violence Against Women*, 20(10), 1179–1202. <https://doi.org/10.1177/1077801214551288>.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>.
- Debnam, K. J., & Mauer, V. (2021). Who, when, how, and why bystanders intervene in physical and psychological teen dating violence. *Trauma, Violence, and Abuse*, 22(1), 54–67. <https://doi.org/10.1177/1524838018806505>.
- DeGue, S., Valle, L. A., Holt, M. K., Massetti, G. M., Matjasko, J. L., & Tharp, A. T. (2014). A systematic review of primary prevention strategies for sexual violence perpetration. *Aggression and Violent Behavior*, 19(4), 346–362. <https://doi.org/10.1016/j.avb.2014.05.004>.

- De La Rue, L., Polanin, J. R., Espelage, D. L., & Pigott, T. D. (2017). A meta-analysis of school-based interventions aimed to prevent or reduce violence in teen dating relationships. *Review of Educational Research*, 87(1), 7–34. <https://doi.org/10.3102/0034654316632061>.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3–4), 327–350. <https://doi.org/10.1007/s10464-008-9165-0>.
- Elliott, M. A., Brewster, S. E., Thomson, J. A., Malcolm, C., & Rasmussen, S. (2015). Testing the bi-dimensional effects of attitudes on behavioral intentions and subsequent behavior. *British Journal of Psychology*, 106(4), 656–674. <https://doi.org/10.1111/bjop.12107>.
- Elliott, M. A., McCartan, R., Brewster, S. E., Coyle, D., Emerson, L., & Gibson, K. (2017). An application of the prototype willingness model to drivers' speeding behavior. *European Journal of Social Psychology*, 47(6), 735–747. <https://doi.org/10.1002/ejsp.2268>.
- Ellsberg, M., Ugarte, W., Ovince, J., Blackwell, A., & Quintanilla, M. (2020). Long-term change in the prevalence of intimate partner violence: A 20-year follow-up study in León, Nicaragua, 1995–2016. *BMJ Global Health*, 5(4). <https://doi.org/10.1136/bmjgh-2020-002339>.
- Eriksen, S. (2015). *Program Evaluation 2015: The Mentors in Violence Prevention Leadership Training at California State University, Long Beach*. <http://www.mvpstrat.com/training-materials/evaluations/differences>.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New-York, NY: Psychology Press.
- Fox, C., Ralph, N., Harrison, E., McElwee, J., Bagnall, C., & Garnett, N. (2020). *Evaluation of the Mentors in Violence Prevention (MVP) program in the West Midlands: Final Report*. Keele University.
- Fox, C., & Vickers, G. (2017). *An evaluation of Mentors in Violence Prevention (MVP) in the West Midlands: The first year of implementation in six Birmingham schools*. Keele University.
- Gibbons, F. X., & Gerrard, M. (1995). Predicting young adults' health risk behavior. *Journal of Personality and Social Psychology*, 69(3), 505–517. <https://doi.org/10.1037/00223514.69.3.505>.
- Gibbons, F. X., & Gerrard, M. (1997). Health images and their effects on health behavior. In B. Bunk & F. X. Gibbons (Eds.), *Health, coping and wellbeing: Perspectives from social comparison theory* (pp. 63–94). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hawkins, D. L., Pepler, D. J., & Craig, W. M. (2001). Naturalistic observations of peer interventions in bullying. *Social Development*, 10(4), 512–527. <https://doi.org/10.1111/14679507.00178>.
- Heisterkamp, A., Fleming, M., & Waitt, C. (2011). *The Sioux City project preventing violence and bullying: A school district and community in motion*. Sioux City, IA: Waitt Institute for Violence Prevention.
- Hester, M., & Westmarland, N. (2005). *Tackling domestic violence: Effective interventions and approaches*. London: Home Office.
- Hindes, S., & Fileborn, B. (2020). “Girl power gone wrong”: #MeToo, Aziz Ansari, and media reporting of (grey area) sexual violence. *Feminist Media Studies*, 20(5), 639–656. <https://doi.org/10.1080/14680777.2019.1606843>.
- Home Office. (2013). *Domestic violence and abuse*. London: Home Office.
- Hoxmeier, J. C., Flay, B. R., & Acock, A. C. (2018). Control, norms, and attitudes: Differences between students who do and do not intervene as bystanders to sexual assault. *Journal of Interpersonal Violence*, 33(15), 2379–2401. <https://doi.org/10.1177/0886260515625503>.
- Hunter, S. C., Russell, K., Pagani, S., Munro, L., Pimenta, S., Marín-López, I., Hong, J. S., & Knifton, L. (2021). A social-ecological approach to understanding adolescent sexting behavior. *Archives of Sexual Behavior*, 50(6), 2347–2357. <https://doi.org/10.1007/s10508-021-01988-9>.
- Jouriles, E. N., Krauss, A., Vu, N. L., Banyard, V. L., & McDonald, R. (2018). Bystander program 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. <https://doi.org/10.1080/07448481.2018.1431906>.
- Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa antibullying program: Grades 4–6. *Child Development*, 82(1), 311–330. <https://doi.org/10.1111/j.1467-8624.2010.01557.x>.
- Katz, J. (1995). Reconstructing masculinity in the locker room: The Mentors in Violence Prevention project. *Harvard Educational Review*, 65(2), 163–175. <https://doi.org/10.17763/haer.65.2.55533188520136u1>.
- Katz, J. (2018). Bystander training as leadership training: Notes on the origins, philosophy, and pedagogy of the mentors in violence prevention model. *Violence Against Women*, 24(15), 1755–1776. <https://doi.org/10.1177/1077801217753322>.
- Katz, J., Heisterkamp, H. A., & Fleming, W. M. (2011). The social justice roots of the Mentors in Violence Prevention model and its application in a high school setting. *Violence Against Women*, 17(6), 684–702. <https://doi.org/10.1177/10778012111409725>.
- Katz, J., & Moore, J. (2013). Bystander education training for campus sexual assault prevention: An initial meta-analysis. *Violence and Victims*, 28(6), 1054–1067. <https://doi.org/10.1891/0886-6708.vv-d-12-00113>.
- Kettrey, H., & Marx, R. (2019). The effects of bystander programs on the prevention of sexual assault across the college years: A systematic review and meta-analysis. *Journal of Youth and Adolescence*, 48(2), 212–227. <https://doi.org/10.1007/s10964-018-0927-1>.
- Koss, M. P., Goodman, L. A., Browne, A., Fitzgerald, L. F., Keita, G. P., & Russo, N. F. (1994). *No safe haven: Male violence against women at home, at work, and in the community*. American Psychological Association. <https://doi.org/10.1037/10156-000>.
- Kovalenko, A. G., Abraham, C., Graham-Rowe, E., Levine, M., & O'Dwyer, S. (2020). What works in violence prevention among young people?: A systematic review of reviews. *Trauma, Violence, and Abuse*. <https://doi.org/10.1177/1524838020939130>.
- Leone, R. M., & Parrott, D. J. (2019). Acute alcohol intoxication inhibits bystander intervention behavior for sexual aggression among men with high intent to help. *Alcoholism: Clinical and Experimental Research*, 43(1), 170–179. <https://doi.org/10.1111/acer.13920>.
- Lester, S., Lawrence, C., & Ward, C. L. (2017). What do we know about preventing school violence? A systematic review of systematic reviews. *Psychology, Health and Medicine*, 22(1), 187–223. <https://doi.org/10.1080/13548506.2017.1282616>.
- Lundgren, R., & Amin, A. (2015). Addressing intimate partner violence and sexual violence among adolescents: Emerging evidence of effectiveness. *Journal of Adolescent Health*, 56(1), 542–550. <https://doi.org/10.1016/j.jadohealth.2014.08.012>.
- Maxwell, C., Chase, E., Warwick, I., Aggleton, P., & Wharf, H. (2010). *Freedom to achieve. Preventing violence, promoting equality: A whole-school approach*. London: Woman-kind Worldwide.
- McCartan, R., & Elliott, M. A. (2018). Bi-dimensional attitudes, attitude accessibility and speeding behavior. *Transportation Research Part F: Traffic Psychology and Behavior*, 58, 581–593. <https://doi.org/10.1016/j.trf.2018.06.036>.
- McCartan, R., Elliott, M. A., Pagani, S., Finnegan, E., & Kelly, S. W. (2018). Testing the effects of explicit and implicit bi-dimensional attitudes on objectively measured speeding behavior. *British Journal of Social Psychology*, 57(3), 630–651. <https://doi.org/10.1111/bjso.12255>.
- McMahon, S., Peterson, N. A., Winter, S. C., Palmer, J. E., Postmus, J. L., & Koenick, R. A. (2015). Predicting bystander behavior to

- prevent sexual assault on college campuses: The role of self-efficacy and intent. *American Journal of Community Psychology*, 56(1–2), 46–56. <https://doi.org/10.1007/s10464-015-9740-0>.
- Midgett, A., Doumas, D., Sears, D., Lundquist, A., & Hausheer, R. (2015). A bystander bullying psychoeducation program with middle school students: A preliminary report. *The Professional Counselor*, 5(4), 486–500. <https://doi.org/10.15241/am.5.4.486>.
- Miller, E., Tancredi, D. J., McCauley, H. L., Decker, M. R., Virata, M. C. D., Anderson, H. A., Stetkevich, N., Brown, E. W., Moideen, F., & Silverman, J. G. (2012). “Coaching boys into men”: A cluster-randomized controlled trial of a dating violence prevention program. *Journal of Adolescent Health*, 51(5), 431–438. <https://doi.org/10.1016/j.jadohealth.2012.01.018>.
- Miller-Halegoua, S. M. (2017). Why null results do not mean no results: negative findings have implications for policy, practice, and research. *Translational Behavioral Medicine*, 7(2), 137–137. <https://doi.org/10.1007/s13142-017-0500-6>.
- Morrison, A., Ellsberg, M., & Bott, S. (2007). Addressing gender-based violence: A critical review of interventions. In S. Devarajan & G. Feder (Eds.) *The World Bank Observer*, 22(1), pp. 25–51. Oxford: Oxford University Press.
- Munafò, M., & Neill, J. (2016). Null is beautiful: on the importance of publishing null results. *Journal of Psychopharmacology*, 30(7), 585–585. <https://doi.org/10.1177/02698811166638>.
- MVP Progress Report (2016). *MVP Progress Report 2015-16*. Education Scotland. <https://education.gov.scot/improvement/practice-exemplars/mentors-for-violence-prevention-mvp-an-overview#>.
- MVP Progress Report (2019). *MVP Progress Report 2018-19*. Education Scotland. <https://education.gov.scot/media/b4c4d0a4/mvp-progress-report-2018-19.pdf>.
- Pagani, S., Hunter, S. C., & Elliott, M. A. (2022a). Bystander intervention among secondary school pupils: Testing an augmented prototype willingness model. *British Journal of Social Psychology*. <https://doi.org/10.1111/bjso.12534>.
- Pagani, S., Hunter, S. C., & Elliott, M. A. (2022b). Evaluating the Mentors in Violence Prevention Program: A process examination of how implementation can affect gender-based violence outcomes. *Journal of Interpersonal Violence*. <https://doi.org/10.1177/08862605221115117>.
- Perkins, H. W., & Berkowitz, A. D. (1986). Perceiving the community norms of alcohol use among students: Some research implications for campus alcohol education programming. *International Journal of the Addictions*, 21(9–10), 961–976. <https://doi.org/10.3109/10826088609077249>.
- Polanin, J. R., Espelage, D. L., & Pigott, T. D. (2012). A meta-analysis of school-based bullying prevention programs’ effects on bystander intervention behavior. *School Psychology Review*, 41(1), 47–65. <https://doi.org/10.1080/02796015.2012.12087375>.
- Rosval, L. (2013). Utilizing the theory of planned behavior to examine the cognitive and social determinants of behavioral responses to bully/victim problems in middle and secondary school students. PhD Thesis, Faculty of Graduate and Postdoctoral Studies, University of Ottawa.
- Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior*, 15(2), 112–120. <https://doi.org/10.1016/j.avb.2009.08.007>.
- Scottish Executive. (2002). *Evaluation of the zero tolerance ‘respect’ pilot project*. Edinburgh, UK: Reid Howie Associates Ltd.
- Sjögren, B., Thornberg, R., Wänström, L., & Gini, G. (2021). Bystander behavior in peer victimisation: Moral disengagement, defender self-efficacy and student-teacher relationship quality. *Research Papers in Education*, 36(5), 588–610. <https://doi.org/10.1080/02671522.2020.1723679>.
- Skipper, Y., & Fox, C. (2021). Boys will be boys: Young people’s perceptions and experiences of gender within education. *Pastoral Care in Education*. <https://doi.org/10.1080/02643944.2021.1977986>.
- Stanley, N., Ellis, J., & Bell, J. (2011). Delivering preventive programs in schools: Identifying gender issues. In C. Barter & D. Berridge (Eds.), *Children behaving badly? Peer violence between children and young people* (pp. 217–230). London, UK: John Wiley and Sons.
- Stanley, N., Ellis, J., Farrelly, N., Hollinghurst, S., & Downe, S. (2015). Preventing domestic abuse for children and young people: A review of school-based interventions. *Children and Youth Services Review*, 59, 120–131. <https://doi.org/10.1016/j.childyouth.2015.10.018>.
- Storer, H., Casey, E., & Herrenkohl, T. (2016). Efficacy of bystander programs to prevent dating abuse among youth and young adults: A review of the literature. *Trauma, Violence, and Abuse*, 17(3), 256–269. <https://doi.org/10.1177/1524838015584361>.
- Thornberg, R., & Jungert, T. (2014). School bullying and the mechanisms of moral disengagement. *Aggressive Behavior*, 40(2), 99–108. <https://doi.org/10.1002/ab.21509>.
- Thornberg, R., & Wänström, L. (2018). Bullying and its association with altruism toward victims, blaming the victims, and classroom prevalence of bystander behaviors: a multilevel analysis. *Social Psychology of Education*, 21(5), 1133–1151. <https://doi.org/10.1007/s11218-018-9457-7>.
- Ward, K. J. (2001). *Mentors in Violence Prevention program evaluation 1999–2000*. Unpublished report, Northeastern University. Boston, MA.
- Williams, D. J., & Neville, F. G. (2013). *Mentors in Violence Prevention: Evaluation of the pilot in Scottish High Schools. End of project report*. University of St Andrews.
- Williams, D. J., & Neville, F. G. (2017). Qualitative evaluation of the mentors in violence prevention pilot in Scottish high schools. *Psychology of Violence*, 7(2), 213–223. <https://doi.org/10.1037/vio0000046>.
- Wilson, C., Woolfson, L. M., Durkin, K., & Elliott, M. A. (2016). The impact of social cognitive and personality factors on teachers’ reported inclusive behavior. *British Journal of Educational Psychology*, 86(3), 461–480. <https://doi.org/10.1111/bjep.12118>.

Stefania Pagani is a Senior Research Officer with the Scottish Government and has a keen interest in interventions designed to reduce gender-based violence.

Simon Hunter is a Professor of Applied Psychology at Glasgow Caledonian University. His main research interest is in youth health and well-being.

David Lawrence is a Professor of Mental Health at Curtin University. His research focuses on the mental health and well-being of both youth and first responders.

Mark Elliott is a Senior Lecturer at the University of Strathclyde. His research seeks to understand decision-making and behavior change and focuses on the socio-cognitive determinants of behavior.