Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

Is the KiVa Anti-bullying Programme an Effective School-based Intervention for Reducing Bullying and Victimisation?

Summary

KiVa is a research based anti-bullying intervention originating in Finland. It is a whole-school intervention, which stems from research on the social standing of aggressive individuals and the participant role approach to bullying. KiVa aims to stop ongoing bullying, prevent new bullying incidents from occurring and reduce the harmful effects of victimisation through changing bystanders' behaviour towards bullying. Until now, there has not been a systematic review of the evidence for KiVa in Finland or worldwide, highlighting a need for this. Thus, a systematic literature review was conducted to evaluate the effectiveness of KiVa on levels of bullying and victimisation. Four databases were searched yielding five studies from Finland, Italy and the United Kingdom (UK). These were assessed for quality and relevance using Gough's (2007) Weight of Evidence Framework. All five studies showed that KiVa reduced levels of bullying and victimisation. However, the results were only consistently significant for ages 7-12 and effect sizes were small. Limitations and future research are discussed.

Introduction

KiVa Anti-bullying Programme

KiVa is a whole-school anti-bullying programme that has been developed at the University of Turku, Finland. KiVa stands for 'Kiusaamista Vastaan' which means 'against bullying' in Finnish (Salmivali, Kärnä, & Poskiparta, 2010). The programme has been developed for children aged 7-15 and has three different versions: Unit 1 (ages 7-9), Unit 2 (ages 10-12) and Unit 3 (ages 13-15). The programme involves universal and indicated actions that aim to stop ongoing bullying, prevent new bullying incidents from developing and decrease the harmful consequences of victimisation (Salmivali, Kärnä, & Poskiparta, 2010).

Universal components

Lessons are used to raise awareness of how the class may influence bullying, they also aim to increase children's empathy towards victims, provide safe strategies for supporting victims and improve victims' coping skills (Haataja et al., 2014; Kärnä et al., 2011a; Salmivali, Kärnä, & Poskiparta, 2010). Throughout the year, classroom teachers carry out 20 hours of lessons involving discussion, role-play, video-clips about bullying, group work and written tasks (Kärnä et al., 2011a). Lessons focus on the topic of bullying; children learn what bullying is, its different forms, consequences and how individuals and groups can reduce it. The lessons also focus on social skills; children learn about emotions, respecting others, being part of a team and group dynamics (Hutchings & Clarkson, 2015; Kärnä et al., 2013; Salmivali, Kärnä, & Poskiparta, 2010).

In addition to lessons, KiVa utilises online activities that are linked with the lesson topics (Salmivali, Kärnä, & Poskiparta, 2010). The objectives of the online components are to provide children with knowledge, incentive and the ability to change their behaviour around bullying (Doll, Pfohl & Yoon, 2010). Units 1 and 2 include an anti-bullying computer game and Unit 3 includes an internet forum 'KiVa Street'.

Universal components also include the use of KiVa materials and parent resources. KiVa posters are displayed around the school and staff are encouraged to wear high visibility vests at break and lunchtimes to remind children to use behaviour that adheres to KiVa principles (Hutchings & Clarkson, 2015). Parents are sent a guide with information about KiVa, bullying and practical advice on how to identify whether their child might be a bully or victim (Nocentini & Menesini, 2016; Salmivali, Kärnä, & Poskiparta, 2010).

Indicated actions

KiVa includes detailed actions that should be carried out by the 'KiVa team' to address confirmed cases of bullying. The team usually includes teachers, other members of school staff and/or EPs (Hutchings & Clarkson, 2015). The team first talks with the victim(s), then the bully(ies) and subsequently carries out a follow up meeting with the victim(s) to establish whether bullying has stopped. In addition, the class teacher meets two to four pro-social and high-status classmates to encourage them to support the victim(s) (Kärnä et al., 2011a; Salmivali, Kärnä, & Poskiparta, 2010).

Psychological Basis

KiVa's theoretical underpinning stems from social-cognitive theory (Bandura, 1989) and views bullying as a social behaviour. Within the social-cognitive framework, the programme is based on two key strands of research; the social standing of aggressive individuals (Cillessen & Mayeux, 2004; Juvonen & Galvan, 2008; South & Wood, 2006) and the participant role approach to bullying (Salmivalli, Lagerspetz, Bjorkqvist, Österman, & Kaukiainen, 1996).

It has been suggested that a desire to achieve high social status within a peer group may contribute to bullying (Salmivali, Kärnä, & Poskiparta, 2010; South & Wood, 2006) and that aggressive behaviour towards peers can facilitate the maintenance of social status (Cillessen & Mayeux, 2004; Juvonen & Galvan 2008). In addition, bullying can be regarded as a group process in which bystanders are not neutral (Salmivalli et al., 1996). Research suggests that bystanders' reactions to bullying incidents can maintain or decrease the bully's behaviour (Salmivalli et al., 1996). If the bystander chooses to reinforce the bully, this can support the bully to achieve high social status, which maintains the bullying behaviour. In contrast, if the bystander chooses to defend the victim, the bully does not attain these social rewards and this decreases the likelihood of bullying behaviour (Salmivalli, Voeten, & Poskiparta, 2011).

KiVa therefore focuses on changing bystanders' attitudes and behaviours towards bullying with a view to decrease the social rewards received through the bullying process and subsequently an individual's desire to bully (Nocentini & Menesini, 2016; Kärnä et al., 2013).

Rationale

Bullying is described as aggressive behaviour that is intentional, conducted repeatedly over time and involves an discrepancy of power (Olweus, 1997). Bullying has been identified as an international problem (Olweus, 2010) which has substantial negative effects on the victim, bully and peers witnessing the incidents. Victims of bullying are likely to experience anxiety, depression, low self-esteem, loneliness and lowered academic achievement (Card & Hodges, 2008; Graham & Juvonen, 1998; Hawker & Boulton, 2000; Isaacs, Hodges, & Salmivalli, 2008). Bullies are also likely to suffer from depression (Salmon, James, & Smith, 1998) and are at greater risk of becoming involved in alcohol abuse and being unemployed later on in life (Kaltiala-Heino, Rimpelä, Rantanen, & Rimpelä, 2000; Kokko & Pulkkinen, 2000). In addition, research suggests that witnessing bullying incidents can have a negative impact on children's mental health (Rivers, Poteat, Noret, & Ashurst, 2009). Therefore, there is a clear need for evidence-based interventions that prevent and reduce levels of bullying and victimisation.

A recent meta-analysis on bullying interventions highlighted that whole-school interventions are the most effective at reducing bullying and victimisation (Ttofi & Farrington, 2011). KiVa is a recently developed anti-bullying programme that utilises a whole-school approach. The programme is used in over 90% of comprehensive schools in Finland (Saarento, Boulton, & Salmivalli, 2014) and is beginning to be implemented around the world (KiVa International, n.d). In particular, the Bangor Centre in the UK is a licenced training centre for KiVa and offers training for Units 1 and 2 (Hutchings & Clarkson, 2016). Since Ttofi and Farrington's (2011) review, studies on the effectiveness of KiVa have emerged globally. KiVa has a promising

evidence base in Finland (Kärnä et al., 2011b) and an emerging evidence base in the Netherlands, Estonia, Italy and the UK (Hutchings & Clarkson, 2015; KiVa International, n.d; Nocentini & Menesini, 2016; Veenstra, 2014).

EPs play a key role in supporting schools to reduce levels of bullying and victimisation. To do this, it is important that EPs recommend interventions that are evidence-based. If KiVa is found to be an effective intervention in reducing bullying and victimisation in schools then this will help to inform EP practice in this area. There is yet to be a systematic review of the evidence for KiVa in Finland or worldwide, highlighting a need for a systematic review of the research in this area.

Review Question

Is the KiVa anti-bullying programme an effective school based intervention for reducing bullying and victimisation?

Critical Review of the Evidence

Literature Search

A comprehensive literature search was conducted on 8th January 2017 using four electronic databases: PsychINFO, Medline, ERIC (EBSCO) and Web of Science. The exact search terms used to locate the studies are presented in Table 1.

Table 1

Search Terms Entered into Databases

1		2		3
KiVa	AND	bully* OR victim*	AND	school* OR child* OR student* OR adolescen*

The initial search yielded 116 studies in total. 62 duplicates were identified and subsequently removed. Following this, the titles and abstracts of 54 studies were screened for suitability using inclusion/exclusion criteria (see Table 2). This left 17 studies, which were screened at full text. During the full text screening, a further 12 studies were removed which left five studies eligible for the review (see Figure 1 for a flow diagram of the study selection process; see Appendix A for the list of removed studies alongside reason for exclusion).
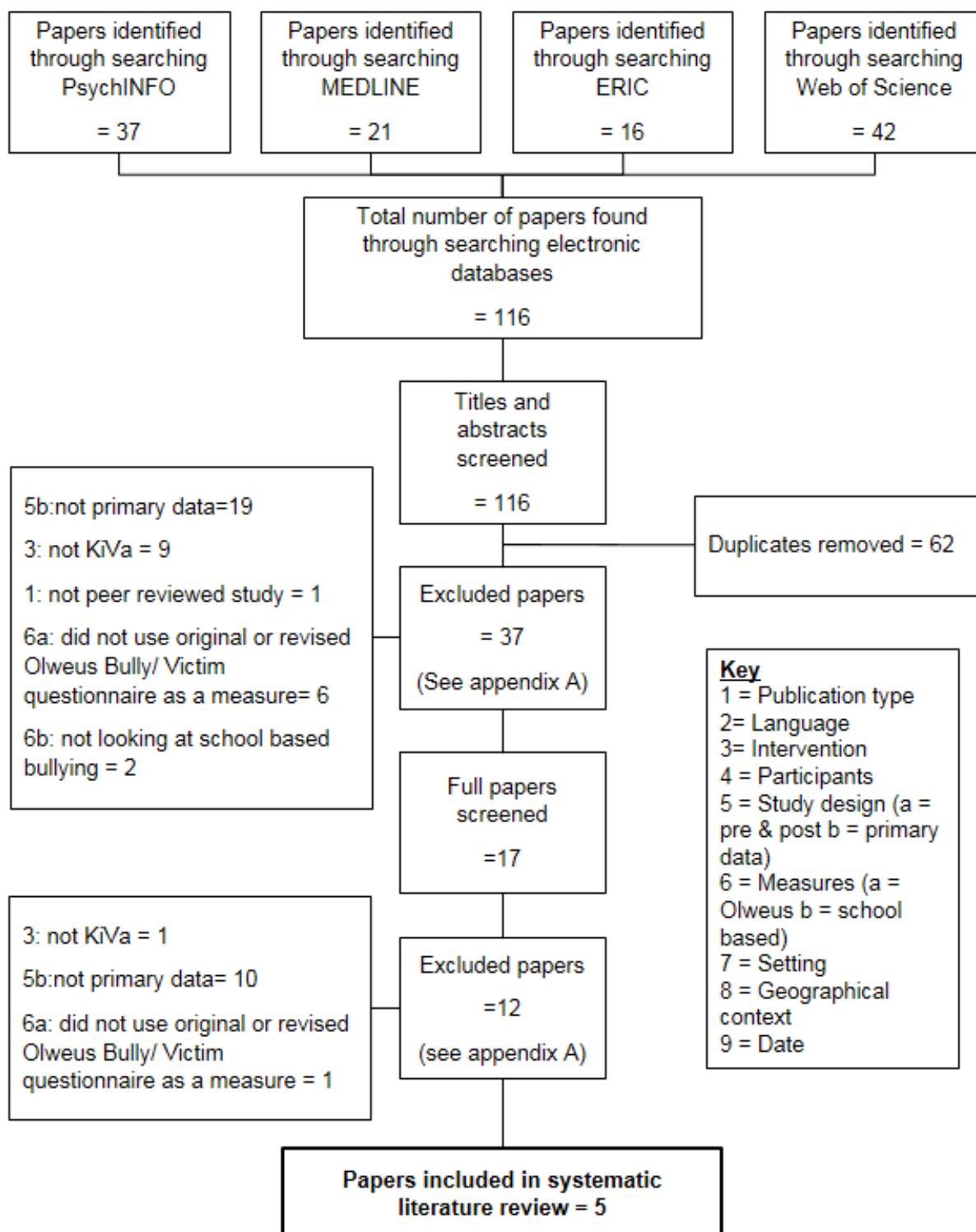
Table 2

Inclusion/Exclusion Criteria

Criterion	Inclusion criteria	Exclusion criteria	Rationale
1. Publication Type	Peer reviewed journal articles.	Not a peer reviewed journal.	Peer reviewed journal articles ensure a level of quality checks have been completed.
2. Language	The study is written in English.	The study is not written in English.	So that the paper can be read and understood the paper needs to be written in English.
3. Intervention	The intervention must be KiVa.	The intervention used is not KiVa.	The review question is specifically looking at the effectiveness of KiVa.

Criterion	Inclusion criteria	Exclusion criteria	Rationale
4. Participants	Children must be aged between 7 – 15 years old.	Children aged younger or older than defined age range.	This is the age range that is targeted by KiVa.
5. Study Design	a) The study design must have pre and post measures. b) The study must contain primary data.	a) The study design does not have pre and post measures. b) The study does not contain primary data.	a) Collecting pre- and post-test data enables the comparison of effects to previous studies using similar assessment methods. b) This will allow the review to examine the effectiveness of the KiVa anti-bullying programme.
6. Measures	a) The study must use the original or revised Olweus Bully/ Victim questionnaire. b) The outcomes must measure school based bullying.	a) The study does not use the original or revised Olweus Bully/ Victim questionnaire. b) The outcomes are not measuring school based bullying (e.g. cyberbullying).	a) The review question is evaluating the effectiveness of KiVa on levels of victimisation and bullying. The original and revised Olweus Bully/Victim questionnaire measures specifically this. It is used by researchers worldwide and has acceptable construct validity and reliability (Kyriakides, Kaloyirou, & Lindsay, 2006). Using the same outcome measure allows comparisons. b) The review question is specifically looking at bullying behaviour within a school setting.
7. Setting	The intervention must be carried out within a school setting.	The intervention is not carried out within a school setting.	The review question is specifically examining the effectiveness of school-based interventions.
8. Geographical Context	Any geographical context.	None.	The review question is looking to review effectiveness of KiVa worldwide as it is being used globally.
9. Date	Studies published on or before 8th January 2017.	Studies published after 8th January 2017.	This was the final search date before the analysis and write up was conducted.

Figure 1

Flow Diagram for Study Selection Process

Papers identified through searching PsychINFO = 37	Papers identified through searching MEDLINE = 21	Papers identified through searching ERIC = 16	Papers identified through searching Web of Science = 42

Total number of papers found through searching electronic databases = 116

Titles and abstracts screened = 116

Duplicates removed = 62

5b:not primary data=19

3: not KiVa = 9

1: not peer reviewed study = 1

6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure= 6

6b: not looking at school based bullying = 2

Excluded papers = 37 (See appendix A)

Full papers screened =17

3: not KiVa = 1

5b:not primary data= 10

6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure = 1

Excluded papers =12 (see appendix A)

Key
1 = Publication type
2= Language
3= Intervention
4 = Participants
5 = Study design (a = pre & post b = primary data)
6 = Measures (a = Olweus b = school based)
7 = Setting
8 = Geographical context
9 = Date

Papers included in systematic literature review = 5

Searches for additional relevant studies were conducted by reviewing reference lists and carrying out a citation search for the five identified studies. Authors with published studies in the area (Judy Hutchings & Susan Clarkson; Hutchings & Clarkson, 2015) were contacted to ensure relevant studies had not been missed. It was noted that a random control trial examining the effectiveness of KiVa for children aged 7-11 was carried out in the UK in 2012. Results from this trial are in the process of being published so were not available to review. The research section of the KiVa anti-bullying website was also searched for additional studies. No additional relevant studies were identified.

Overall, five studies were deemed eligible for critical review (see Table 3). A summary table, which includes detailed characteristics of each study, can be found in Appendix B.

Table 3

List of Studies Identified as Suitable for Critical Review

Eligible Studies
1　Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011a). A Large-Scale Evaluation of the KiVa Antibullying Program: Grades 4-6. *Child Development*, *82*(1), 311–330.
2　Nocentini, A., & Menesini, E. (2016). KiVa Anti-Bullying Program in Italy: Evidence of Effectiveness in a Randomized Control Trial. *Prevention Science*, *17*(8), 1012–1023.
3　Kärnä, A., Voeten, M., Little, T. D., Alanen, E., Poskiparta, E., & Salmivalli, C. (2013). Effectiveness of the KiVa Antibullying Program: Grades 1–3 and 7–9. *Journal of Educational Psychology*, *105*(2), 535–551.
4　Hutchings, J., & Clarkson, S. (2015). Introducing and piloting the KiVa bullying prevention programme in the UK. *Educational and Child Psychology*, *32*(1), 49–61.
5　Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Alanen, E., & Salmivalli, C. (2011b). Going to scale: A nonrandomized nationwide trial of the KiVa antibullying program for grades 1–9. *Journal of Consulting and Clinical Psychology*, *79*(6), 796–805.

Evaluation Process

Weight of Evidence

The five studies were assessed for quality and relevance using Gough's (2007) Weight of Evidence (WoE) framework. This framework can be used to appraise studies on four key dimensions: methodological quality (WoE A), methodological relevance (WoE B), topic relevance (WoE C) and overall assessment (WoE D).

WoE A of each study was determined using the adapted version of Kratochwill's (2003) APA Task Force coding protocol for group designs. Adaptions made to Kratochwill's (2003) protocol are outlined in Appendix C, and one example of a completed coding protocol is detailed in Appendix D. Criteria for appraising WoE B were created based on the Guyatt et al. (1995) hierarchy of evidence. Criteria for WoE C were based on how relevant each study was to the review question (details on criteria and the rationale for all WoE ratings are outlined in Appendix E). The weightings for each area were averaged to produce WoE D. Table 4 provides a summary of the weightings given to each study and Table 5 provides the classification of scores.

Table 4

Summary of Weight of Evidence Judgements

Study	WoE A Methodological Quality	WoE B Methodological Relevance	WoE C Topic Relevance	WoE D Overall weight of evidence
Kärnä et al. (2011a)	1.5 Medium	2 Medium	3 High	2.17 Medium
Nocentini & Mensini (2016)	1.38 Low	2 Medium	2 Medium	1.79 Medium
Kärnä et al. (2013)	1.5 Medium	2 Medium	3 High	2.17 Medium
Hutchings & Clarkson (2015)	.88 Low	1 Low	2 Medium	1.29 Low

Kärnä et al. (2011b)	1.25 Low	2 Medium	3 High	2.08 Medium

Table 5

Classification of Scores

WoE D Overall weight of evidence	Averaged Scores
High	≥ 2.5
Medium	1.4 - 2.5
Low	≤ 1.4

Participants

The five studies chosen for review include data from 331,839 pupils ranging from 7-15 years-old. Three studies looked into the effectiveness of KiVa for pupils aged 9-12[1] (Hutchings & Clarkson, 2015; Kärnä et al., 2011a; Nocentini & Mensini, 2016), one study looked at pupils aged 7-9 and 12-15 (Kärnä et al., 2013) and one study looked at pupils aged 7-15 (Kärnä et al., 2011b). Sample sizes varied across studies, with one study including 297,728 pupils (Kärnä et al., 2011b) and one including 473 (Hutchings & Clarkson, 2015). All sample sizes were adequate for the statistical analysis carried out.

Three studies were conducted in Finland (Kärnä et al., 2011a; Kärnä et al.,2011b; Kärnä et al., 2013), one in the UK (Hutchings & Clarkson, 2015) and one in Italy (Nocentini & Mensini, 2016). The studies judged to be the highest quality came from Finland, followed by Italy then the UK, scoring 'medium', 'medium' and 'low' respectively.

---

[1] Hutchings and Clarkson (2015) looked at pupils aged 9-11 but used Unit 2, which was developed for pupils aged 10-12. Kärnä et al. (2011a) and Nocentini & Mensini (2016) looked at pupils aged 10-12.

The percentage of males and females was reported for four studies (Hutchings & Clarkson, 2015; Kärnä et al., 2011a; Kärnä et al.,2011b; Nocentini & Mensini, 2016) and was roughly equal. Kärnä et al. (2013) did not explicitly reference the percentage of males and females in their study but accounted for and evaluated gender differences in the analysis. The three Finish studies noted that they considered their sample representative of the intended population (Finnish comprehensive schools). The UK study provided limited information regarding pupil demographics and the Italian study noted that children were not necessarily representative of the population (see Appendix B for detailed characteristics of each study). Information regarding demographics was reflected in the scores given in WoE C.

Intervention Details

Three studies implemented the original version of KiVa (Kärnä et al., 2011a; Kärnä et al., 2011b; Kärnä et al., 2013) and one study adapted the language but kept all elements of the intervention the same (Hutchings & Clarkson, 2015). One study adapted the language alongside making changes to pictures and stories, substituting vests with badges and excluding the online components (Nocentini & Mensini, 2016). Although the core components of KiVa were kept the same, it was thought that this version did not accurately reflect the KiVa intervention and only scored a 'medium' on WoE C because of this.

Fidelity was considered an important factor for the evaluation of the studies in WoE A. If schools are provided with high quality training, support and supervision to carry out the programme, then this will ensure that the intervention is carried out as

intended. Four studies noted that they had ongoing supervision alongside training manuals and training days (Hutchings & Clarkson, 2015; Kärnä et al., 2011a; Kärnä et al., 2013; Nocentini & Mensini, 2016). One study did not make reference to any ongoing supervision (Kärnä et al., 2011b), which is reflected in its 'low' score for WoE A.

Outcome Measures

All five studies utilised the revised Owleus Bully/Victim Questionnaire or key global questions from the Owleus Bully/Victim Questionnaire to measure bullying and victimisation. This measure is deemed a valid and reliable measure of bullying and victimisation (Kyriakides et al., 2006). However, use of this measure independently, without any additonal measures was deemed a weakness in the methodology section of WoE A, as self-report measures can be affected by social-desirability bias (Barker, Pistrang & Elliott, 2016). The use of additional measures utilising different methods or sources allows triangulation of data and therefore reduces bias. Three studies only used self-report methods (Hutchings & Clarkson, 2015; Kärnä et al., 2011b; Nocentini & Mensini, 2016). Two studies used both peer and self-report methods (Kärnä et al., 2011a; Kärnä et al., 2013). None of the studies utilised data collection from multiple sources, which prevented any of the studies from achieving a weighting of 'high' for WoE A.

Study Design

In line with evidence hierarchies (Guyatt et al., 1995), randomised control trials were judged to be the highest quality for WoE B. 'No intervention' control groups withhold potentially beneficial interventions from a group, therefore an 'active' control group

design was deemed more ethical and thus higher quality for WoE A. Three studies used randomised control trial study designs with 'no intervention' control groups (Kärnä et al., 2011a; Kärnä et al., 2013; Nocentini & Mensini, 2016). One study made use of a non-randomised design (cohort longitudinal design with adjacent cohorts; Kärnä et al., 2011b). The design ensured that group equivalence was established through post-hoc analysis, which enabled it to score a 'medium' on WoE B. One study did not have a control group (Hutchings & Clarkson, 2015). Studies without control groups risk having confounding variables influence bullying and victimisation scores, therefore the study design was not deemed high quality and scored 'low' on WoE A and B.

None of the studies included in the review incorporated a follow-up assessment. Absence of follow-up assessments within the studies reviewed means that they can only report the short-term effects of KiVa, over one year. This weakness is reflected in the scores for WoE A.

Findings

Table 6 summarises the results and corresponding effect sizes for the primary outcomes of all five studies. Positive effects of KiVa have been noted in all five studies. However, the effectiveness appears to be mediated by age and some gender differences were noted. Effect sizes ranged from 'not practically significant' to 'large'.

For ages 7-9 (Unit 1), Kärnä et al. (2011b) and Kärnä et al. (2013) found a significant decrease in self-reported victimisation and bullying. The effect sizes for significant

findings consistently favoured the intervention but were classed as 'not practically significant' as they were below .2 (Cohen, 1992).

For ages 10-12[2] (Unit 2), Kärnä et al. (2011a), Kärnä et al. (2011b), Nocentini and Mensini (2016) and Hutchings and Clarkson (2015) all found a significant decrease in self-reported victimisation and bullying. In addition, Kärnä et al. (2011a) found that peer-reported levels of victimisation significantly decreased, but peer-reported levels of bullying did not. The effect sizes for these findings varied between 'not practically significant' to 'large'. Kärnä et al. (2011a) found a 'small' effect size for peer-reported victimisation and 'not practically significant' effect sizes for all other significant findings. Kärnä et al. (2011b) found 'not practically significant' effect sizes for all significant outcomes. In the Italian school system, grades 4 and 5 are in Elementary School and grade 6 is in Middle School. Nocentini and Mensini (2016) found 'not practically significant' effect sizes on key global questions from the Olweus Bully/Victim Questionnaire and 'small' effect sizes from the Florence Bully-Victimization scale for both Primary and Middle Schools. The effects sizes were largest in Elementary Schools. Hutchings and Clarkson (2015) found a 'medium' effect size for self-reported victimisation and a 'large' effect size for self-reported bullying. However, caution should be exercised when interpreting these results as the study was evaluated as having a 'low' weighting. In particular, the lack of a control group may have biased the results.

For ages 13-15 (Unit 3), Kärnä et al. (2013) found a significant decrease in peer-reported victimisation, but not for self-reported victimisation and bullying or peer-

---

[2] Hutchings and Clarkson (2015) looked at pupils aged 9-11 but used Unit 2, which was developed for pupils aged 10-12. Kärnä et al. (2011a) and Nocentini & Mensini (2016) looked at pupils aged 10-12.

reported bullying. The effect size found was 'not practically significant'. Kärnä et al. (2011b) did not find a significant decrease in self-reported victimisation or bullying.

The majority of the effect sizes yielded from the studies reviewed were small or too weak to be considered meaningful. However, these effect sizes fit in with the current literature on the effectiveness of anti-bullying interventions. Ttofi and Farrington's (2011) meta-analysis highlighted that the average odds ratio (OR) for anti-bullying programmes was 1.29 (95% CI [1.17,1.41]) for victimisation and 1.36 (95% CI [1.26, 1.47]) for bullying. The corresponding standardised mean difference (*d*) values were .14 for bullying and .17 for victimisation. In line with Cohen (1992) and Ferguson (2009) these would be considered 'not practically significant'. This suggests that the effect sizes yielded by KiVa are a similar size to other anti-bullying interventions.

Table 6

Summary of Results and Effect Size

Study	Participant numbers	Primary outcomes	Outcome measure	Analysis	Measure of effect size and calculation	Effect size	Descriptor[a]	WoE D
Kärnä et al. (2011a) Study ID: 1	Total sample size: 8166 Intervention group: 4201 Control group: 3965	Self- and peer-reported victimisation (*b* = -0.154, p <.001, *b* = -0.085, p <.05) and self-reported bullying (*b* = -0.309, p <.001) significantly decreased for pupils in KiVa schools, compared to control schools. Peer-reported bullying did not significantly decrease (*b*= -0.130, p = ns).	*Self-reported victimisation* Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996) *Self-reported bullying* Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996) *Peer-reported victimisation* Participant role Questionnaire (Salmivalli et al., 1996) *Peer-reported bullying* Participant role Questionnaire (Salmivalli et al., 1996)	Multilevel regression	Extracted from the article. Cohen's *d* was calculated by dividing the adjusted group mean difference by the pooled within-group standard deviation (Morris, 2008). Extracted from the article. Odds ratios (OR) were calculated using LOR logarithm (Ttofi & Farrington, 2011).	*Self-reported victimisation* At 4 months: *d* = 0.03 At 9 months: *d* = 0.17 *Self-reported bullying* At 4 months: *d* = 0.05 At 9 months: *d* = 0.10 *Peer-reported victimisation* At 4 months: *d* = 0.18 At 9 months: *d* = 0.33 *Peer-reported bullying* At 4 months: *d* = 0.10 At 9 months: *d* = 0.14 *Victimisation (9 months)* OR = 1.47 (95% CI [1.10-1.96]) *Bullying (9 months)* OR = 1.22 (95% CI [0.78-1.90])	NPS[b] NPS NPS NPS NPS Small NPS NPS NPS NPS	Medium

Study	Participant numbers	Primary outcomes	Outcome measure	Analysis	Measure of effect size and calculation	Effect size	Descriptor[a]	WoE D
Nocentini & Mensini (2016) Study ID: 2	Total sample size: 2042 Intervention group: 1039 Control group: 1003	<u>Elementary School</u> Victimisation and bullying significantly decreased for pupils in KiVa schools, compared to control schools (B = -0.027, p <.001, B = -0.012, p <.05). <u>Middle School</u> Victimisation and bullying significantly decreased for pupils in KiVa schools, compared to control schools (B = -0.023, p <.001, B = -0.013, p <.001).	*Self-reported victimisation* The Florence Bullying-Victimization Scale (Palladino et al., 2016) *Self-reported bullying* The Florence Bullying-Victimization Scale (Palladino et al., 2016) *Self-reported victimisation* Key global questions from Olweus Bully/ Victim Questionnaire (Olweus, 1996) *Self-reported bullying* Key global questions from Olweus Bully/ Victim Questionnaire (Olweus, 1996)	Linear mixed-effect model	Extracted from the article. Cohen's d was calculated by dividing the adjusted group mean difference by the pooled within-group standard deviation (Morris, 2008). For this measure, the mean and standard deviation were not published so it was not possible to calculate Cohen's $d.$ Odds ratios (OR) were extracted from the article and calculated using LOR logarithm (Ttofi & Farrington, 2011).	*Self-reported victimisation (9 months)* Elementary school: d = 0.38 Middle school: d = 0.24 *Self-reported bullying* Elementary school: d = 0.23 Middle school: d = 0.21 *Self-reported victimisation* Elementary school: OR =1.93 (95% CI [1.744-2.125]) Middle school: OR = 1.22 (95% CI [0.938 – 1.576]) *Self-reported bullying* Elementary school: OR =1.31 (95% CI [0.930 – 2.125]) Middle school: OR = 1.33 (95% CI [0.927 – 1.892])	Small Small Small Small NPS NPS NPS NPS	Medium

Study	Participant numbers	Primary outcomes	Outcome measure	Analysis	Measure of effect size and calculation	Effect size	Descriptor[a]	WoE D
Kärnä et al. (2013) Study ID: 3	Total sample size: 23430 Intervention group: 76 schools Control group: 71 schools	<u>Grades 2– 3</u> Self- reported victimisation and bullying significantly decreased for pupils in KiVa schools, compared to control schools (b = -0.49, p <.01, b = -0.36, p <.05). <u>Grades 8 – 9</u> Peer reported victimisation decreased significantly for pupils in KiVa schools, compared to control schools (b = -0.10, p <.001). Self- reported victimisation (b = -0.04, p = ns) and peer and self -reported bullying (b = 0, p = ns, b = -0.08, p = ns) did not decrease significantly compared to control schools.	*Self-reported victimisation* Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996) *Self-reported bullying* Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996) *Peer-reported victimisation* Participant role Questionnaire (Salmivalli et al., 1996) *Peer-reported bullying* Participant role Questionnaire (Salmivalli et al., 1996)	Multilevel regression	Cohen's d was calculated by dividing the adjusted group mean difference by the pooled within-group standard deviation (Morris, 2008). Extracted from	*Self-reported victimisation* Grades 2-3 At 4 months: d = 0.06 At 9 months: d = 0.09 Grades 8-9 At 4 months: d = 0.03 At 9 months: d = 0.03 *Self-reported bullying* Grades 2-3 At 4 months: d = 0.04 At 9 months: d = 0.08 Grades 8-9 At 4 months: d = 0.04 At 9 months: d = 0.04 *Peer-reported victimisation* Grades 8-9 At 4 months: d = 0.01 At 9 months: d = 0.01 *Peer-reported bullying* Grades 8-9 At 4 months: d = 0 At 9 months: d = 0 *Victimisation (9 months)* Grades 2-3 OR = 1.34 (95% CI [1.07-1.61]) *Grades 8-9* OR = 1.04	 NPS NPS NPS NPS NPS NPS NPS NPS NPS NPS No effect No effect NPS	Medium

Study	Participant numbers	Primary outcomes	Outcome measure	Analysis	Measure of effect size and calculation	Effect size	Descriptor[a]	WoE D
					the article. Odds ratios (OR) were calculated using LOR logarithm (Ttofi & Farrington, 2011).	(95% CI [0.86-1.22]) *Bullying (9 months)* Grades 2-3 OR = 1.43 (95% CI [1.10-1.77]) *Grades 8-9* OR = 1.08 (95% CI [0.88-1.28])	NPS NPS NPS	
Hutchings & Clarkson (2015) Study ID: 4	Total sample size: 473 No control group	Self- reported victimisation and bullying significantly decreased (*t* = 2.147, p <.05, *t* = 2.791, p <.01).	*Self-reported victimisation* Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996) *Self-reported bullying* Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996)	Repeated measures *t*-test	Cohen's *d* was calculated by dividing the difference between pre and post means by the pre-test standard deviation score (Becker, 1988).	*Self-reported victimisation* *d* = 0.62 *Self-reported bullying* *d* =1.41	Medium Large	Low
Kärnä et al. (2011b) Study ID: 5	Total sample size: 297728 Intervention group: 156634 Control	<u>Grades 1-6</u> Victimisation and bullying significantly decreased (ICC = 0.06, p <.001, ICC = 0.09, p <.001). Intervention effects increased from Grade 1 until Grade 4. After Grade 4 the	*Self-reported victimisation* Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996)	Intraclass correlations	For this measure, the mean and standard deviation were not published so it was not possible to calculate Cohen's *d.* Odds ratios (OR) were	*Self-reported victimisation* Grades 1-9 OR = 1.22 (95% CI [1.19, 1.24]) Grades 1-3 OR = 1.21 (95% CI [1,11, 1.32]) Grades 4-6 OR = 1.28 (95% CI [1.16, 1.42])	Medium NPS NPS NPS	Medium

Study	Participant numbers	Primary outcomes	Outcome measure	Analysis	Measure of effect size and calculation	Effect size	Descriptor[a]	WoE D
	group: 141103	effects decreased. <u>Grades 7-9</u> Victimisation and bullying did not significantly decrease. (ICC = 0.03, p = ns, ICC = 0.04, p = ns). Effects only reached statistical significance for victimisation in Grade 8.There was a larger drop in programme effectiveness for bullying than for victimisation.	*Self-reported bullying* Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996)		extracted from the article and calculated using LOR logarithm (Ttofi & Farrington, 2011).	Grades 7-9 OR = 1.13 (95% CI [0.98, 1.30]) *Self-reported bullying* Grades 1-9 OR = 1.18 (95% CI [1.15, 1.21]) Grades 1-3 OR = 1.24 (95% CI [1.11, 1.40]) Grades 4-6 OR = 1.25 (95% CI [1.08, 1.45]) Grades 7-9 OR = 1.06 (95% CI [0.91, 1,22])	NPS NPS NPS NPS NPS	

[a]Cohen's *d* interpretations based on 0.2 = 'small' effect size, 0.5 = 'medium' effect size and 0.8 = 'large' effect size (Cohen, 1992). Odds ratio interpretations based on 2.0 = 'recommended minimum' effect size, 3.0 = 'moderate' effect size and 4.0 = 'strong' effect size (Ferguson, 2009).
[b]Not practically significant (NPS).
[c]No data presented for peer-reported victimisation or bullying for grades 2-3.

Section 4: Conclusions and Recommendations

Conclusions

The purpose of this review was to evaluate the effectiveness of KiVa for reducing levels of bullying and victimisation. The five studies included in the review were deemed to have weight of evidence scores ranging from 'low' to 'medium'. Four studies were rated as 'medium' and achieved effect sizes that ranged from 'not practically significant' to 'small'. One study rated as 'low' achieved effect sizes that ranged from 'medium' to 'large'.

All five studies noted that KiVa reduced levels of bullying and victimisation. However, the results were only consistently significant for ages 7-12 (Unit 1 and 2) and effect sizes were largest for ages 10-12 (Unit 2). This suggests that the intervention is most effective in children up to age 12, which adds to the mixed literature around age and anti-bullying interventions. It fits in with Smith's (2010) findings that anti-bullying interventions have a lager impact on younger children, and contradicts Ttofi and Farrington's (2011) findings that anti-bullying interventions have a larger impact on older children. Reasons for KiVa's differential effect on age may be due to individual factors; as the child gets older, bullying behaviour is thought to become more stable and therefore harder to influence (Monks, Smith, & Swettenham, 2005). Organisational factors have also been proposed as reasons why anti-bullying interventions are more effective for younger children (Smith, 2010). However, from this review there is insufficient evidence to look at organisational factors as most of the studies were conducted in Finland where children attend the same school from age 7-16.

Findings suggest that KiVa can reduce bullying and victimisation levels in Finland and other countries; significant results were found in Finland, Italy and the UK. Studies from Finland were deemed the highest quality, followed by Italy then the UK. The UK's pilot study showed promising results, but due to weaknesses in the methodology, results should be interpreted with caution.

In conclusion, although KiVa can be seen to decrease levels of bullying and victimisation in children aged 7-12, it is likely that this change is small and in some cases too small to be considered meaningful. This is in line with the effectiveness of other anti-bullying interventions and raises questions around the impact of anti-bullying interventions. In relation to practice, EPs should be recommending interventions that are evidence-based. The evidence from this review highlights that KiVa can reduce bullying and victimisation in children aged 7-12. However, EPs should be mindful that KiVa may only produce small effects and consider the cost effectiveness of implementing a whole-school intervention.

Recommendations for Future Research

It was not within the scope of this review to evaluate the cost-effectiveness of KiVa. Implementing interventions that are cost effective is of importance to schools and policy makers, which highlights the need for a cost-benefit analysis of KiVa.

In addition, the results from the studies reviewed all relied solely on questionnaire data, which can be affected by social desirability bias and pose a threat to the validity of the results obtained (Barker, Pistrang & Elliott, 2016). Furthermore, none of the studies collected follow up data so it was not possible to determine whether

KiVa is effective in the long term. Future research should therefore use outcome measures which utilise multi-method techniques and study designs which incorporate a follow-up assessment.

Three of the five studies were conducted by the same research team (Kärnä et al.), which includes four researchers who are from the University where KiVa was developed. This research team may have an investment in endorsing the effectiveness of KiVa, which could possibly bias conclusions drawn across these studies. It could be beneficial for further research to be carried out by different research teams in order to remove this potential bias.

Finally, the review only looked at a small number of studies, five, which means that generalisations should be made with caution. In particular, there was only one study from the UK, which was deemed to have 'low' weighting. In order to make confident generalisations of KiVa's effectiveness in the UK, high quality studies are needed. Publication of the recent random control trial in the UK may provide this.

References

Bandura, A. (1989). Social cognitive theory. In *Annals of child development: Six theories of child development* (Vol. 6, pp. 1–60). Greenwich, CT: JAI Press.

Barker, C., Pistrang, N., & Elliott, R. (2016). *Research methods in clinical psychology an introduction for students and practitioners* (3rd ed.). Chichester: Wiley-Blackwell.

Becker, B. J. (1988). Synthesizing standardized mean change measures. *British Journal of Mathematical and Statistical Psychology*, *41*, 257–278.

Card, N. A., & Hodges, E. V. E. (2008). Peer victimization among schoolchildren: Correlations, causes, consequences, and considerations in assessment and intervention. *School Psychology Quarterly*, *23*(4), 451–461.

Cillessen, A. H., & Mayeux, L. (2004). From Censure to Reinforcement: Developmental Changes in the Association Between Aggression and Social Status. *Child Development*, *75*, 147–163.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Doll, B., Pfohl, W., & Yoon., J. S. (2010). *Handbook of Youth Prevention Science.* (pp. 238–252). Routledge/ Taylor & Francis Group.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538.

Graham, S., & Juvonen, J. (1998). Self-blame and peer victimization in middle school: an attributional analysis. *Developmental Psychology*, *34*(3), 587–599.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213-228.

Guyatt, G. Sackett, M., Sinclair, J., Hayward, R., Cook, D., & Cook, R. (1995). User's guides to medical literature. A Method for Grading Health Care Recommendations. *Journal of the American Medical Association*, (2754), 1800–1804.

Haataja, A., Voeten, M., Boulton, A. J., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2014). The KiVa antibullying curriculum and outcome: Does fidelity matter? *Journal of School Psychology*, *52*(5), 479–493.

Hawker, D. S., & Boulton, M. J. (2000). Twenty years' research on peer victimization and psychosocial maladjustment: a meta-analytic review of cross-sectional studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *41*(4), 441–455.

Hutchings, J., & Clarkson, S. (2015). Introducing and piloting the KiVa bullying prevention programme in the UK. *Educational and Child Psychology*, *32*(1), 49–61.

Hutchings, J., & Clarkson, S. (2016). KiVa Anti-bullying Programme. Gwynedd: Bangor University.

Isaacs, J., Hodges, E. V. E., & Salmivalli, C. (2008). Long-term Consequences of Victimization by Peers: A Follow-up from Adolescence to Young Adulthood. *European Journal of Developmental Science*, *2*(4), 387–397.

Juvonen, J., & Galvan, A. (2008). Peer influence in involuntary social groups: Lessons from research on bullying. In *Understanding peer influence in children and adolescents* (pp. 225–244). New York, NY: Guilford Press.

Kaltiala-Heino, R., Rimpelä, M., Rantanen, P., & Rimpelä, A. (2000). Bullying at school--an indicator of adolescents at risk for mental disorders. *Journal of Adolescence*, *23*(6), 661–74.

Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011a). A Large-Scale Evaluation of the KiVa Antibullying Program: Grades 4-6. *Child Development*, *82*(1), 311–330.

Kärnä, A., Voeten, M., Little, T. D., Alanen, E., Poskiparta, E., & Salmivalli, C. (2013). Effectiveness of the KiVa Antibullying Program: Grades 1–3 and 7–9. *Journal of Educational Psychology*, *105*(2), 535–551

Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Alanen, E., & Salmivalli, C. (2011b). Going to scale: A nonrandomized nationwide trial of the KiVa antibullying program for grades 1–9. *Journal of Consulting and Clinical Psychology*, *79*(6), 796–805.

Kokko, K., & Pulkkinen, L. (2000). Aggression in childhood and long-term unemployment in adulthood: a cycle of maladaptation and some protective factors. *Developmental Psychology*, *36*(4), 463–472.

Kratochwill, T. R. (2003). Task Force on Evidence-Based Interventions in School Psychology. *American Psychology Association.*

Kyriakides, L., Kaloyirou, C., & Lindsay, G. (2006). An analysis of the Revised Olweus Bully/Victim Questionnaire using the Rasch measurement model. *The British Journal of Educational Psychology*, *76*, 781–801.

KiVa International. (n.d). Evidence of effectiveness in Finland and elsewhere. Retrieved January 8, 2017, from http://www.kivaprogram.net/is-kiva-effective

Menesini, E., Codecasa, E., Benelli, B., & Cowie, H. (2003). Enhancing children's responsibility to take action against bullying. Evaluation of a befriending intervention in Italian middle schools. *Aggressive Behavior,* (29),1–14.

Monks, C. P., Smith, P. K., & Swettenham, J. (2005). Psychological correlates of peer victimisation in preschool: Social cognitive skills, executive function and attachment profiles. *Aggressive Behavior*, *31*(6), 571–588.

Morris, S. B. (2008). Group Designs. *Organizational Research Methods*, (April 2003), 364–386.

Nocentini, A., & Menesini, E. (2016). KiVa Anti-Bullying Program in Italy: Evidence of Effectiveness in a Randomized Control Trial. *Prevention Science*, *17*(8), 1012–1023.

Olweus, D. (1997). Bully/victim problems in school. *European Journal of Psychology of Education*, *7*(4), 495–510.

Olweus, D. (2010). Understanding and researching bullying some critical issues. In *Handbook of bullying in schools: An international perspective* (pp.9–33). Routledge/ Taylor & Francis Group: US.

Olweus, D. (1996). *The revised Olweus bully/victim questionnaire*. University of Bergen, Research Center for Health Promotion.

Palladino, B. E., Nocentini, A., & Menesini, E. (2016). Evidence-based intervention against bullying and cyberbullying: Evaluation of the NoTrap! program in two independent trials. *Aggressive behavior*, *42*(2), 194-206.

Pöyhönen, V., Kärnä, A.,& Salmivalli, C. (2008). Symposium conducted at the 20th Biennial Meeting of the International Society for the Study of Behavioural Development. Germany: Würzburg.

Pöyhönen, V., Juvonen, J., & Salmivalli, C. (2010). What does it take to stand up for the victim of bullying? The interplay between personal and social factors. *Merrill-Palmer Quarterly,* 56, 143–163.

Rigby, K., & Slee, P. T. (1991). Bullying among Australian school children: Reported behavior and attitudes toward victims. *Journal of Social Psychology*, 131, 615–627.

Rivers, I., Poteat, V. P., Noret, N., & Ashurst, N. (2009). Observing bullying at school: The mental health implications of witness status. *American Psychological Association*, *24*(4), 211–223.

Saarento, S., Boulton, A. J., & Salmivalli, C. (2014). Reducing Bullying and Victimization: Student- and Classroom-Level Mechanisms of Change. *Journal of Abnormal Child Psychology*, 1–16.

Salmivalli, C., Karna, A., & Poskiparta, E. (2010). From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program. In *Handbook of bullying in schools: An international perspective.* (pp. 441-454). Routledge/ Taylor & Francis Group: US.

Salmivalli, C., Lagerspetz, K., Bjorkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a Group Process: Particant Roles and Their Relations to Social Status Within the Group. *Aggressive Behavior*, *22*, 1–15.

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.

Salmon, G., James, a, & Smith, D. M. (1998). Bullying in schools: self reported anxiety, depression, and self esteem in secondary school children. *BMJ*, *317*, 924–925.

Smith, P. K. (2010). Bullying in primary and secondary schools: psychological and organizational comparisons. In *Handbook of bullying in schools: An international perspective* (pp. 137–150). New York: Routledge.

South, C. R., & Wood, J. (2006). Bullying in Prisons: The Importance of PErceived Social Status, Prisonization, and Moral Disengagement. *Aggressive Behaviour*, *32*, 490–501.

Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, *7*(1), 27–56.

Veenstra, R. (2014). Groepsprocessen bij jongeren: over pesten en ander probleemgedrag. *Kind en Adolescent*, *35*, 86-99

Appendices

Appendix A: Excluded Studies.

Appendix B: Summary of Studies.

Appendix C: Adaptions to Kratochwill's (2003) APA Task Force Coding Protocol for Group Designs.

Appendix D: Example of a Completed Coding Protocol.

Appendix E: Details on Criteria and the Rationale for all WoE Ratings.

Appendix A

Excluded Studies

Excluded paper: Abstract Search	Rationale for Exclusion
Adams, R., Taylor, J., Duncan, A., & Bishop, S. (2016). Peer Victimization and Educational Outcomes in Mainstreamed Adolescents with Autism Spectrum Disorder (ASD). *Journal of Autism and Developmental Disorders*, *46*(11), 3557–3566.	3: not KiVa.
Ahtola, A., Haataja, A., Karna, A., Poskiparta, E., & Salmivalli, C. (2013). Implementation of anti-bullying lessons in primary classrooms: How important is head teacher support? *Educational Research*,*55*(4), 376–392.	6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure.
Ahtola, A., Haataja, A., Kärnä, A., Poskiparta, E., & Salmivalli, C. (2012). For children only? Effects of the KiVa antibullying program on teachers. *Teaching and Teacher Education*, *28*(6), 851–859.	6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure.
Bradshaw, C. P. (2015). Translating research to practice in bullying prevention. *The American Psychologist*, *70*(4), 322–32.	5b: not primary data.
Chaux, E., Velasquez, A. M., Schultze-Krumbholz, A., & Scheithauer, H. (2016). Effects of the cyberbullying prevention program media heroes (Medienhelden) on traditional bullying. *Aggressive Behavior*, *42*(2), 157–165.	3: not KiVa.
Clarkson, S., Axford, N., Berry, V., Edwards, R. T., Bjornstad, G., Wrigley, Z., & Hutchings, J. (2016). Effectiveness and micro- costing of the KiVa school-based bullying prevention programme in Wales: study protocol for a pragmatic definitive parallel group cluster randomised controlled trial. *BMC Public Health*, *16*(1), 104.	5b: not primary data.
Costable, A., & Spears, B. *(2012). The Impact of Technology on Relationships in Educational Settings.* Routledge/ Taylor & Francis Group*.*	3: not KiVa.
Doll, B., Pfohl, W., & Yoon., J. S. (2010). *Handbook of Youth Prevention Science.* (pp. 484). Routledge/ Taylor & Francis Group.	3: not KiVa.
Elledge, C. L., Williford, A., Boulton, A. J., DePaolis, K. J., Little, T. D., & Salmivalli, C. (2013). Individual and Contextual Predictors of Cyberbullying: The Influence of Children's Provictim	6b: not looking at school based bullying.

Excluded paper: Abstract Search	Rationale for Exclusion
Attitudes and Teachers' Ability to Intervene. *Journal of Youth and Adolescence*, *42*(5), 698–710.	
Fox, K. A., & Shjarback, J. A. (2016). What Works to Reduce Victimization? Synthesizing What We Know and Where to Go From Here. *Violence and Victims*, *31*(2), 285–319.	5b: not primary data.
Garandeau, C. F., Poskiparta, E., & Salmivalli, C. (2014). Tackling acute cases of school bullying in the KiVa anti-bullying program: A comparison of two approaches. *Journal of Abnormal Child Psychology*, *42*(6), 981–991.	6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure.
Haataja, A., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2015). A process view on implementing an antibullying curriculum: How teachers differ and what explains the variation. *School Psychology Quarterly*, *30*(4), 564–576.	6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure.
Kim, S. K., & Kim, N. S. (2013). The role of the pediatrician in youth violence prevention. *Korean Journal of Pediatrics*, *56*(1), 1–7.	3: not KiVa.
Mayes, S. D., Baweja, R., Calhoun, S. L., Syed, E., Mahr, F., & Siddiqui, F. (2014). Suicide ideation and attempts and bullying in children and adolescents: Psychiatric and general population samples. *Crisis*, *35*(5), 301–309.	3: not KiVa.
Menzer, M. M., & Torney-Purta, J. (2012). Individualism and socioeconomic diversity at school as related to perceptions of the frequency of peer aggression in fifteen countries. *Journal of Adolescence*, *35*(5), 1285–1294.	3: not KiVa
Nocentini, A., Zambuto, V., & Menesini, E. (2015). Anti-bullying programs and Information and Communication Technologies (ICTs): A systematic review. *Aggression and Violent Behavior*, *23*, 52–60.	5b: not primary data
Noland, B. (2011). *Effects of the KiVa Anti-Bullying Program on Adolescents' Perception of Peers, Depression, and Anxiety*. Universiry of Kansas.	1: not peer reviewed study.
Olweus, D. (2013). School Bullying: Development and Some Important Challenges. In *Annual Review of Clinical Psychology*. (pp. 751-780).	5b: not primary data

Excluded paper: Abstract Search	Rationale for Exclusion
Ostrov, J. M., & Kamper, K. E. (2015). Future Directions for Research on the Development of Relational and Physical Peer Victimization. *Journal of Clinical Child & Adolescent Psychology*, *4416*, 1–11.	5b: not primary data
Pecorini, A., Nocentini, A., & Menesini, E. (2016). A Systemtic Review on Virtual Environment Projects to Prevent School Bullying. *Psicologia Clinica Dello Sviluppo*, 20(1), 27-54.	3: not KiVa.
Poskiparta, E., & Salmivalli, C. (2012). Kiva Antibullying program: Overview of Education Studies Based on a Randomized Controlled Trial and National Rollout in Finland. *International Journal of Conflict and Violence*, *6*(2), 294–302.	5b: not primary data
Ryan, A., & Ladd, G. W. (2012). *Peer relationships and adjustment at school.* (pp. 279-305). IAP Information Age Publishing; US.	5b: not primary data.
Saarento, S., Garandeau, C. F., & Salmivalli, C. (2015). Classroom- and School-Level Contributions to Bullying and Victimization: A Review. *Journal of Community & Applied Social Psychology*, *25*, 204–218.	5b: not primary data
Saarento, S., & Salmivalli, C. (2015). The Role of Classroom Peer Ecology and Bystanders' Responses in Bullying. *Child Development Perspectives*, *9*(4), 201–205.	5b: not primary data.
Saarento, S., Boulton, A. J., & Salmivalli, C. (2014). Reducing Bullying and Victimization: Student- and Classroom-Level Mechanisms of Change: A Review. *Journal of Abnormal Child Psychology*, 1–16.	5b: not primary data.
Salmivalli, C., & Poskiparta, E. (2012). Making bullying prevention a priority in Finnish schools: The KiVa antibullying program. In *Evidence-based bullying prevention programs for children and youth.* (pp. 41 – 53). Jossey-Bass; US.	5b: not primary data.
Salmivalli, C., & Poyhonen, V. (2012) Cyberbullying in Finland. In *Cyberbullying in the global playground: Research from the international perspectives.* (pp. 57 – 72). Wiley-Blackwell.	5b: not primary data.
Salmivalli, C., Garandeau, C. F., & Veenstra, R. (2012). KiVa anti-bullying program:	5b: not primary data.

Excluded paper: Abstract Search	Rationale for Exclusion
Implications for school adjustment. In *Peer relationships and adjustment at school.* (pp.279-305). IAP Information Age Publishing; US.	
Salmivalli, C., Karna, A., & Poskiparta, E. (2010). From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program. In *Handbook of bullying in schools: An international perspective.* (pp. 441-454). Routledge/ Taylor & Francis Group: US.	5b: not primary data.
Salmivalli, C., Peets, K., & Hodges, E. V. E. (2011). Bullying. In *The Wiley-Blackwell Handbook of Childhood Social Development Second Edition.* (pp. 510– 528). Blackwell Publishing.	5b: not primary data.
Salmivalli, C., Poskiparta, E., Ahtola, A., & Haataja, A. (2013). The implementation and effectiveness of the KiVa antibullying program in Finland. *European Psychologist*, *18*(2), 79–88.	5b: not primary data.
Sentse, M., Veenstra, R., Kiuru, N., & Salmivalli, C. (2015). A Longitudinal Multilevel study of Individual Characteristics and Classroom Norms in Explaining Bullying Behaviors. *Journal of Abnormal Child Psychology*, *43*(5), 943–955.	6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure.
Smith, P. K., Salmivalli, C., & Cowie, H. (2012). Effectiveness of school-based programs to reduce bullying: A commentary. *Journal of Experimental Criminology*, *8*(4), 433–441.	5b: not primary data
Stronhmeir, D., & Noam, G. G. (2012). *Evidence-based bullying prevention programs for children and youth*. Jossey-Bass; US.	5b: not primary data.
Veenstra, R., Verlinden, M., Huitsing, G., Verhulst, F. C., & Tiemeier, H. (2013). Behind bullying and defending: Same-sex and other-sex relations and their associations with acceptance and rejection. *Aggressive Behavior*, *39*(6), 462–471.	3: not KiVa.
Williford, A., Elledge, L. C., Boulton, A. J., DePaolis, K. J., Little, T. D., & Salmivalli, C. (2013). Effects of the KiVa Antibullying Program on Cyberbullying and Cybervictimization Frequency Among Finnish Youth. *Journal of Clinical Child & Adolescent Psychology*, *42*(6), 820–833.	6b: not looking at school based bullying.

Excluded paper: Abstract Search	Rationale for Exclusion
Williford, A., Boulton, A., Noland, B., Little, T. D., Kärnä, A., & Salmivalli, C. (2012). Effects of the KiVa anti-bullying program on adolescents' depression, anxiety, and perception of peers. *Journal of Abnormal Child Psychology*, *40*(2), 289–300.	6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure.

Excluded paper: Full paper search	Rationale for Exclusion
Garandeau, C. F., Lee, I. A., & Salmivalli, C. (2014). Differential effects of the KiVa anti-bullying program on popular and unpopular bullies. *Journal of Applied Developmental Psychology*, *35*(1), 44–50.	5b: not primary data. Data from (Kärnä et al., 2011).
Haataja, A., Sainio, M., Turtonen, M., & Salmivalli, C. (2015). Implementing the KiVa antibullying program: recognition of stable victims. *Educational Psychology*, *3410*(January), 1–17.	6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure.
Haataja, A., Voeten, M., Boulton, A. J., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2014). The KiVa antibullying curriculum and outcome: Does fidelity matter? *Journal of School Psychology*, *52*(5), 479–493.	5b: not primary data. Data from (Kärnä et al., 2011a & 2013).
Juvonen, J., Schacter, H. L., Sainio, M., & Salmivalli, C. (2016). Can a School-Wide Bullying Prevention Program Improve the Plight of Victims? Evidence for Risk ɤ Intervention Effects. *Journal of Consulting and Clinical Psychology*, *84*(4), 334–344.	5b: not primary data. Data from (Kärnä et al., 2011a).
Saarento, S., Boulton, A. J., & Salmivalli, C. (2014). Reducing Bullying and Victimization: Student- and Classroom-Level Mechanisms of Change. *Journal of Abnormal Child Psychology*, 1–16.	5b: not primary data. Data from (Kärnä et al., 2011a).
Sainio, M., Veenstra, R., Huitsing, G., & Salmivalli, C. (2012). Same- and Other-Sex Victimization: Are the Risk Factors Similar? *Aggressive Behavior*, *38*(6), 442–455.	5b: not primary data. Data from (Kärnä et al., 2011a & 2013).
Salmivalli, C., Karna, a., & Poskiparta, E. (2011). Counteracting	5b: not primary data. Data from (Kärnä et al., 2011a).

Excluded paper: Full paper search	Rationale for Exclusion
bullying in Finland: The KiVa program and its effects on different forms of being bullied. *International Journal of Behavioral Development*, *35*(5), 405–411.	
Sentse, M., Kiuru, N., Veenstra, R., & Salmivalli, C. (2014). A social network approach to the interplay between adolescents' bullying and likeability over time. *Journal of Youth and Adolescence*, *43*(9), 1409–1420.	5b: not primary data. Data from (Kärnä et al., 2011a).
Strohmeier, D., Kärnä, A., & Salmivalli, C. (2011). Intrapersonal and interpersonal risk factors for peer victimization in immigrant youth in Finland. *Developmental Psychology*, *47*(1), 248–258.	5b: not primary data. Data from (Kärnä et al., 2011).
Veenstra, R., Lindenberg, S., Huitsing, G., Sainio, M., & Salmivalli, C. (2014). The Role of Teachers in Bullying: The Relation Between Antibullying Attitudes, Efficacy, and Efforts to Reduce Bullying. *Journal of Educational Psychology*, *106*(4), 1135–1143.	6a: did not use original or revised Olweus Bully/ Victim questionnaire as a measure.
Yang, A., & Salmivalli, C. (2015). Effectiveness of the KiVa Antibullying Programme on Bully-Victims, Bullies and Victims. *Educational Research*, *57*(1), 80–90.	5b: not primary data. Data from (Kärnä et al., 2011).

Appendix B

Summary of Studies

Study	Study design	Sample characteristics [a]	Intervention details	Outcome measures	Duration	Country	Key findings
Kärnä et al. (2011a) Study ID: 1	Random assignment to intervention or control condition at the school level.	Total sample size: 8166 pupils across 70 schools Intervention: 4201 pupils, 39 schools Control: 3965 pupils, 32 schools Age: 9-12 years (Grades 4-6) Gender: 50.1% girls, 49.9% boys Ethnicity: Most students native Finns (Caucasian), 2.4% immigrants Setting Comprehensive schools Considered representative of Finnish comprehensive schools	KiVa: Unit 2	Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996) Participant role Questionnaire (Salmivalli et al., 1996) 20-item Provictim scale (Rigby & Slee, 1991) 7-item empathy scale (Pöyhönen, Kärnä & Salmivalli, 2008) Self-efficacy for defending scale (Poyhonen et al., 2010) Wellbeing scale (Metsamuuronen & Svedlin, 2004)	1 school year (August – May)	Country: Finland Language of instruction: Swedish and Finnish	Self- and peer- reported victimisation and self-reported bullying significantly decreased for pupils in KiVa schools, compared to control schools. Peer- reported levels of assisting and reinforcing the bully significantly decreased for pupils in KiVa schools, compared to control schools. Self- reported self-efficacy for defending and wellbeing at school significantly increased for pupils in KiVa schools, compared to controls.
Nocentini & Mensini (2016) Study ID: 2	Random assignment to intervention or control condition at the school level.	Total sample size: 2042 pupils across 13 schools Intervention: 1039 pupils, 7 schools Control: 1003 pupils, 6 schools Age: 9-12 years (Grades 4-6) Gender: 51% girls, 49% boys Ethnicity: 92% from Italian background SES: 10 schools = medium, 3 schools – low	KiVa: Unit 2 (Adapted version: changes made to language, pictures and stories; vests substituted with	The Florence Bullying-Victimization Scale (Palladino et al., 2016) Key questions from Olweus Bully/ Victim Questionnaire (Olweus, 1996) Revised version of the Questionnaire on Attitudes towards bullying (Menesini et al., 2003)	1 school year (September – June)	Country: Italy Language of instruction: Italian	Elementary School Victimisation and bullying significantly decreased for pupils in KiVa schools, compared to control schools. Pro-victim attitudes and empathy towards the victim significantly increased for pupils in KiVa schools, compared to control schools. Middle School

37

Study	Study design	Sample characteristics [a]	Intervention details	Outcome measures	Duration	Country	Key findings
		Setting Elementary and Middle Schools Not considered representative of Italian population: all schools for control and intervention volunteered, does not account for very low risk or very high risk schools	badges; online components were excluded)	7-item empathy scale (Pöyhönen, Kärnä & Salmivalli, 2008)			Victimisation and bullying significantly decreased for pupils in KiVa schools, compared to control schools. Pro-victim attitudes significantly increased for pupils in KiVa schools, compared to control schools.
Kärnä et al. (2013) Study ID: 3	Random assignment to intervention or control condition at the school level.	Total sample size: 23430 pupils across 147 schools *Grades 1- 3*[b] <u>Total sample size:</u> 6927 across 74 schools Intervention: 38 schools Control: 36 schools <u>Age:</u> 7-9 years <u>Setting:</u> Elementary school *Grades 7 -9* <u>Total sample size:</u> 16503 across 74 schools Intervention: 38 schools Control: 35 schools <u>Age:</u> 13-15 years <u>Setting:</u> Lower Secondary Considered representative of Finnish schools	KiVa: Unit 1 and 3 Within the indicated actions, schools were randomly assigned to (a) confronting approach and (b) a non-confronting approach	Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996) Participant role Questionnaire (Salmivalli et al., 1996)	1 school year (August – May)	<u>Country:</u> Finland <u>Language of instruction:</u> Swedish and Finnish	<u>Grades 1-9</u> Comparison of means revealed an overall decrease in levels of bullying and victimisation. <u>Grades 2– 3</u> [c] Self- reported victimisation and bullying significantly decreased for pupils in KiVa schools, compared to control schools. <u>Grades 8 – 9</u> Peer reported- victimisation decreased significantly for pupils in KiVa schools, compared to control schools. Self- reported victimisation and peer and self -reported bullying did not decrease significantly compared to control schools.
Hutching s &	Quasi-experiment	<u>Total sample size:</u> 473 pupils across 13 schools	KiVa: Unit 2 Version	Revised Olweus Bully/ Victim Questionnaire (Olweus,	1 school year	<u>Country:</u> England	Self- reported victimisation and bullying significantly decreased.

Study	Study design	Sample characteristics [a]	Intervention details	Outcome measures	Duration	Country	Key findings
Clarkson (2015) Study ID: 4	al design: one group pre-test post-test design.	Age: 9-11 years (Years 5-6) Gender: 48.2% girls, 51.8% boys Setting Primary school	translated into English	1996) Teachers reported on the experience of delivering the programme on an online survey	(September – July)	Language of instruction: English	Gender differences: girls showed reductions in victimisation and bullying but boys only showed reductions in bullying. Teachers reported high levels of pupil acceptance and engagement with lessons.
Kärnä et al. (2011b) Study ID: 5	Quasi-experiment al design: Cohort longitudinal design with adjacent cohorts	Total sample size: 297,728 pupils across 888 schools Intervention: 156,634 pupils Control: 141, 103 pupils Age: 7- 15 years (Grades 1-9) Gender: 49% girls, 51% boys Ethnicity: Most students native Finns (Caucasian), 3% immigrants Setting Elementary schools (70.8%), Lower secondary schools (13.3%), both Elementary and Lower secondary grade schools (15.9%) Considered representative of Finnish schools	KiVa: Unit 1, 2 and 3	Revised Olweus Bully/ Victim Questionnaire (Olweus, 1996)	1 school year (August – May)	Country: Finland Language of instruction: Swedish and Finnish	KiVa programme is more effective in Grades 1-6 than in Grades 7-9. Grades 1-6 Victimisation and bullying significantly decreased. Intervention effects increased from Grade 1 until Grade 4. After Grade 4 the effects decreased. Grades 7-9 Victimisation and bullying did not significantly decrease. Effects only reached statistical significance for victimisation in Grade 8.There was a larger drop in programme effectiveness for bullying than for victimisation.

[a]Sample characteristics calculated at time of analysis.

[b]Only collected post-test measures for Grades 1 and 7 as they were not in the schools at the time of pre-test. Grades 2-3: 4704 students (74 schools). Grades 8-9: 11070 students (73 schools).

[c]Multilevel   regression   analysis   only   included   Grades   2-3   and   8-9.

Appendix C

Adaptions to Kratochwill's (2003) APA Task Force Coding Protocol for Group Designs

Section Excluded	Rationale
I. B7. Coding for qualitative research methods.	All studies are using quantitative data.
II. C. Primary/ Secondary Outcomes Are Statistically Significant.	This will be evaluated in other sections of the review through written commentary, tabulation of key findings, effect sizes and WoE D.
II. D. Educational/ Clinical Significance.	Participants were not from a clinical sample. Educational significant is being evaluated as part of the main review.
II. G Replication.	Not relevant to the review question.
II. H. Site of Implementation.	Not relevant as the review is only looking at school-based interventions.
II. A2. Participant characteristics specified for treatment and control groups.	Information already gathered and provided in summary table (Appendix B).
II. A4.Receptivity/ acceptance by target participant population.	Not relevant to the review question, just looking at effectiveness.

Appendix D

Example of a Completed Coding Protocol

Procedural and Coding Manual 45

Coding Protocol: Group-Based Design

Domain:

☐ School- and community-based intervention programs for social and behavioral problems

☐ Academic intervention programs

☐ Family and parent intervention programs

☐ School-wide and classroom-based programs

☐ Comprehensive and coordinated school health services

Name of Coder(s):   Date: 29/01/17

Full Study Reference in APA format: Kama, 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.

Intervention Name (description from study): KiVa Antibullying Program

Study ID Number (Unique Identifier): 1

Type of Publication: (Check one)

☐ Book/Monograph
☐ Journal article
☐ Book chapter
☐ Other (specify):

41

I. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)

A1.1 ☐ Completely randomized design
A1.2 ☐ ==Randomized block design (between-subjects variation)== Blocked by schools: stratified by province and language.
A1.3 ☐ Randomized block design (within-subjects variation)
A1.4 ☐ Randomized hierarchical design

A2. Nonrandomized designs (if nonrandom assignment design, select one of the following)

A2.1 ☐ Nonrandomized design
A2.2 ☐ Nonrandomized block design (between-participants variation)
A2.3 ☐ Nonrandomized block design (within-participants variation)
A2.4 ☐ Nonrandomized hierarchical design
A2.5 ☐ Optional coding of Quasi-experimental designs (see Appendix C)

A3. Overall confidence of judgment on how participants were assigned (select one of the following)

A3.1 ☐ Very low (little basis)
A3.2 ☐ Low (guess)
A3.3 ☐ Moderate (weak inference)
A3.4 ☐ High (strong inference)
A3.5 ☐ ==Very high (explicitly stated)==
A3.6 ☐ N/A
A3.7 ☐ Unknown/unable to code

B. Statistical Treatment/Data Analysis (answer B1 through B6)

B1. Appropriate unit of analysis   ☐ ==yes== ☐ no hierarchical model used
B2. Familywise error rate controlled   ☐ ==yes== ☐ no   ☐ N/A
B3. Sufficiently large N ☐ ==yes== ☐ no
         Statistical Test: multilevel regression analysis
         α level: 0.05
         ES: Medium
         N required: 67

B4. Total size of sample (start of the study): 8237 (dropped to 8166 at analysis)

B5. Intervention group sample size: 4207 (4201 at analysis)

B6. Control group sample size: 4030 (3965 at analysis)

For studies using qualitative research methods, code B7 and B8

B7. Coding

B7.1 Coding scheme linked to study's theoretical-empirical basis (select one) ☐ yes ☐ no

B7.2 Procedures for ensuring consistency of coding are used (select one) ☐ yes   ☐ no
Describe procedures: _____

B7.3 Progression from abstract concepts to empirical exemplars is clearly articulated (select one)   ☐ yes   ☐ no

B8.   Interactive process followed (select one)   ☐ yes   ☐ no

       Describe process: _____

C. Type of Program (select one)

C1. ☐ <mark>Universal prevention program</mark> with indicated actions for cases of bullying which arise within the school
C2. ☐ Selective prevention program
C3. ☐ Targeted prevention program
C4. ☐ Intervention/Treatment
C5. ☐ Unknown

D. Stage of the Program (select one)

D1. ☐ <mark>Model/demonstration programs</mark> first evaluation if KiVa
D2. ☐ Early stage programs
D3. ☐ Established/institutionalized programs
D4. ☐ Unknown

E. Concurrent or Historical Intervention Exposure (select one)

E1. ☐ Current exposure
E2. ☐ Prior exposure
E3. ☐ <mark>Unknown</mark> not stated

II.   Key Features for Coding Studies and Rating Level of Evidence/ Support
(3=Strong Evidence   2=Promising Evidence   1=Weak Evidence   0=No Evidence)

A. Measurement (answer A1 through A4)

A1. Use of outcome measures that produce reliable scores for the majority of primary outcomes. The table for Primary/Secondary Outcomes Statistically Significant allows for listing separate outcomes and will facilitate decision making regarding measurement (select one of the following)

    A1.1 ☐ Yes
    A1.2 ☐ No
    A1.3 ☐ Unknown/unable to code

Revised Olweus Bully/ Victim Questionnaire: α = .85 (not stated in paper but paper made reference to paper where scale has been evaluated). Participant Role Questionnaire: bully scale α = .91, assistant scale α = .90, reinforce scale α =.85, defender scale α = .91, victim scale α = .84. 20-item Provictim scale: α = .79. Seven-item empathy scale: α = .84. Self-efficacy for defending scale: α =.69. Well-being scale: α = .88.

    A2. Multi-method (select one of the following)
    A2.1 ☐ Yes
    A2.2 ☐ No
    A2.3 ☐ N/A
    A2.4 ☐ Unknown/unable to code

Peer and self-report questionnaires.

    A3. Multi-source (select one of the following)
    A3.1 ☐ Yes
    A3.2 ☐ No
    A3.3 ☐ N/A
    A3.4 ☐ Unknown/unable to code

Peer and self- report.

    A4. Validity of measures reported (select one of the following)
    A5.1 ☐ Yes validated with specific target group
    A5.2 ☐ In part, validated for general population only
    A5.3 ☐ No
    A5.4 ☐ Unknown/unable to code

Rating for Measurement (select 0, 1, 2, or 3):   ☐ 3 ☐ 2 ☐ 1 ☐ 0

B. Comparison Group

    B1. Type of Comparison Group (select one of the following)

        B1.1 ☐ Typical contact
        B1.2 ☐ Typical contact (other) specify:
        B1.3 ☐ Attention placebo
        B1.4 ☐ Intervention elements placebo
        B1.5 ☐ Alternative intervention
        B1.6 ☐ PharmacotherapyB1.1
        B1.7 ☐ No intervention
        B1.8 ☐ Wait list/delayed intervention
        B1.9 ☐ Minimal contact
        B1.10 ☐ Unable to identify comparison group

Rating for Comparison Group (select 0, 1, 2, or 3):   ☐ 3 ☐ 2 ☐ 1 ☐ 0

B2. Overall confidence rating in judgment of type of comparison group (select one of the following)

    B2.1 ☐ Very low (little basis)
    B2.2 ☐ Low (guess)
    B2.3 ☐ Moderate (weak inference)
    B2.4 ☐ High (strong inference)
    B2.5 ☐ ==Very high (explicitly stated)==
    B2.6 ☐ Unknown/Unable to code

Schools stratified by province and language then randomly assigned.

B3. Counterbalancing of Change Agents (answer B3.1 to B3.3)

    B3.1 ☐ By change agent
    B3.2 ☐ Statistical
    B3.3 ☐ ==Other==

Not stated, intervention received by school staff in each school following KiVa training.

B4. Group Equivalence Established (select one of the following)

    B4.1 ☐ ==Random assignment==
    B4.2 ☐ Posthoc matched set
    B4.3 ☐ Statistical matching
    B4.4 ☐ Post hoc test for group equivalence

B5. Equivalent Mortality (answer B5.1 through B5.3)

    B5.1 ☐ ==Low Attrition== (less than 20% for Post)
    B5.2 ☐ Low Attrition (less than 30% for follow-up)
    B5.3 ☐ Intent to intervene analysis carried out
           Findings _____

Attrition from study: started at 8237- dropped to 7564 due to consent – 1 whole school dropped out before study – throughout study 251 students left schools and 436 joined schools – from W1 to W2 2 control schools dropped out (51 students) – from W2 to W3 5 controls schools dropped out (640 students) - missing data inputted using SAS Proc MI – conducted 100 imputations using the Markov Chain Monte Carlo algorithm- final sample size 8166 (Intervention = 4201, Control = 3965). Before missing data inputted, attrition rate = 19%.

~~C. Primary/Secondary Outcomes Are Statistically Significant~~

~~C1. Evidence of appropriate statistical analysis for primary outcomes (answer C1.1 through C1.3)~~

    ~~C1.1 ☐ Appropriate unit of analysis (rate from previous code)~~
    ~~C1.2 ☐ Familywise/experimentwise error rate controlled when applicable (rate from previous code) C1.3 ☐ Sufficiently large N (rate from previous code)~~

~~C2. Percentage of primary outcomes that are significant (select one of the following)~~

    ~~C2.1 ☐ Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct~~

    ~~C2.2 ☐ Significant primary outcomes for between 50% and 74% of the total primary outcome measures for each key construct~~

    ~~C2.3 ☐ Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct~~

~~**Rating for Primary Outcomes Statistically Significant (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ 0**~~

~~C3. Evidence of appropriate statistical analysis for secondary outcomes (answer C3.1 through C3.3)~~

    ~~C3.1 ☐ Appropriate unit of analysis~~
    ~~C3.2 ☐ Familywise/experimentwise error rate controlled when applicable (rate from previous code)~~

C3.3 ☐ Sufficiently large N (rate from previous code)

C4. Percentage of secondary outcomes that are significant (select one of the following)

C4.1 ☐

C4.2 ☐   Significant secondary outcomes for at least 75% of the total secondary outcome measures for each key construct

☐ Significant secondary outcomes for between 50% and 74% of the total secondary outcome measures for each key construct

C4.3 ☐ Significant secondary outcomes for between 25% and 49% of the total secondary outcome measures for any key construct

Rating for Secondary Outcomes Statistically Significant (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ 0

C5. Overall Summary of Questions Investigated

C5.1 Main effect analyses conducted   (select one) ☐ yes ☐ no
C5.2 Moderator effect analyses conducted   (select one) ☐ yes ☐ no
Specify results: _____
C5.3 Mediator analyses conducted   (select one) ☐ yes ☐ no
Specify results: _____

C. Primary/Secondary Outcomes Statistically Significant (only list p ≤ .05)

(list primary outcomes first in alphabetical order, followed by secondary outcomes in alphabetical order)

Outcomes	Primary vs. Secondary	Who Changed	What Changed	Source	Treatment Information	Outcome Measure Used	Reliability	ES	(1—)
Outcome #1:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown					
Outcome #2	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown					
Outcome #3:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown					
Outcome #4:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown					
Outcome #5:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown					

Null Findings/Negative Outcomes Associated with the Intervention (listed alphabetically by outcome)

Outcomes	Primary vs. Secondary	Who Was Targeted for Change	What Was Targeted for Change	Source	Rate null/negative outcomes	Outcome Measure Used	Reliability	ES
Outcome #1:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown				
Outcome #2	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown				
Outcome #3:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown				
Outcome #4:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown				
Outcome #5:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown	☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other ☐ Unknown				

Type of Data Effect Size is Based On	Confidence Rating in ES Computation
(check all that apply)	(select one of the following)
☐ Means and SDs ☐ t-value or F – value ☐ Chi-square (df = 1) ☐ Frequencies or proportions (dichotomous) ☐ Frequencies or proportions (polytomous) ☐ Other (specify): ☐ Unknown	☐ Highly estimated (e.g., only have M p value) ☐ Moderate estimation (e.g., have complex but complete statistics) ☐ Some estimation (e.g., unconventional statistics that require conversion) ☐ Slight estimation (e.g., use significance testing statistics rather than descriptives) ☐ No estimation (e.g., all descriptive data is present)

D. Educational/Clinical Significance

Outcome Variables:	Pretest	Posttest	Follow-Up
D1. Categorical Diagnosic Data	Diagnostic information regarding inclusion into the study presented: ☐ Yes ☐ No ☐ Unknown	Positive change in diagnostic criteria from pre to posttest: ☐ Yes ☐ No ☐ Unknown	Positive change in diagnostic criteria from posttest to follow-up: ☐ Yes ☐ No ☐ Unknown
D2. Outcome Accessed via continuous Variables		Positive change in percentage of participants showing clinical improvement from pre to posttest: ☐ Yes ☐ No ☐ Unknown	Positive change in percentage of participants showing clinical improvement from posttest to follow up: ☐ Yes ☐ No ☐ Unknown
D3. Subjective Evaluation: The importance of behavior change is evaluated by individuals in direct contact with the participant.	Importance of behavior change is evaluated: ☐ Yes ☐ No ☐ Unknown	Importance of behavior change from pre to posttest is evaluated positively by individuals in direct contact with the participant ☐ Yes ☐ No ☐ Unknown	Importance of behavior change from posttest to follow up is evaluated positively by individuals in direct contact with the participant ☐ Yes ☐ No ☐ Unknown
D4. Social Comparison: Behavior of participant at pre, post, and follow up is compared to normative data (e.g., a typical peer).	Participant's behavior is compared to normative data ☐ Yes ☐ No ☐ Unknown	Participant's behavior has improved from pre to posttest when compared to normative data: ☐ Yes ☐ No ☐ Unknown	Participant's behavior has improved from posttest to follow up when compared to normative data: ☐ Yes ☐ No ☐ Unknown

Rating for Educational/Clinical Significance (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ 0

E. Identifiable Components (answer E1 through E7)
    E1. Evidence for primary outcomes (rate from previous code): ☐ 3 ☐ 2 ☐ 1 ☐ **0**
    E2. Design allows for analysis of identifiable components (select one) ☐ yes ☐ **no**

E3. Total number of components: 5
(Universal components: student lessons, computer game, KiVa symbols/ posters around school, parent guide. Indicated actions: KiVa team responding to bullying incident). Although the main component is the student lessons.

49

E4. Number of components linked to primary outcomes: 5

<u>Additional criteria to code descriptively:</u>

E5. Clear documentation of essential components (select one) ☐ |yes☐ ☐ no

E6. Procedures for adapting the intervention are described in detail (select one)☐ yes ☐ no

E7. Contextual features of the intervention are documented (select one) ☐ |yes☐ ☐ no

Rating for Identifiable Components (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ 0

F. Implementation Fidelity

    F1. Evidence of Acceptable Adherence (answer F1.1 through F1.3)

        F1.1 ☐ Ongoing supervision/consultation
        F1.2 ☐ Coding intervention sessions/lessons or procedures
        F1.3 ☐ Audio/video tape implementation (select F1.3.1 or F1.3.2):

            F1.3.1 ☐ Entire intervention
            F1.3.2 ☐ Part of intervention

    F2. Manualization (select all that apply)

        F2.1 ☐ Written material involving a detailed account of the exact procedures and
            the sequence in which they are to be used

        F2.2 ☐ Formal training session that includes a detailed account of the
            exact procedures and the sequence in which they are to be used

        F2.3 ☐ Written material involving an overview of broad principles and a description
            of the intervention phases

        F2.4 ☐ Formal or informal training session involving an overview of broad
            principles and a description of the intervention phases

    F3. Adaptation procedures are specified (select one) ☐ yes ☐ no ☐ unknown

Rating for Implementation Fidelity (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ 0

~~G. Replication (answer G1, G2, G3, and G4)~~

    ~~G1. ☐ Same Intervention~~
    ~~G2. ☐ Same Target Problem~~
    ~~G3. ☐ Independent evaluation~~

~~Rating for Replication (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ 0~~

~~H. Site of Implementation~~
  ~~H1. School (if school is the site, select one of the following options)~~

    ~~H1.1 ☐ Public~~

H1.2 ☐ Private
H1.3 ☐ Charter
H1.4 ☐ University Affiliated
H1.5 ☐ Alternative
H1.6 ☐ Not specified/unknown

H2. Non School Site (if it is a non school site, select one of the following options)
H2.1 ☐ Home
H2.2 ☐ University Clinic
H2.3 ☐ Summer Program
H2.4 ☐ Outpatient Hospital
H2.5 ☐ Partial inpatient/day Intervention Program
H2.6 ☐ Inpatient Hospital
H2.7 ☐ Private Practice
H2.8 ☐ Mental Health Center
H2.9 ☐ Residential Treatment Facility
H2.10 ☐ Other (specify):_____
H2.11 ☐ Unknown/insufficient information provided

Rating for Site of Implementation (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ 0

I. Follow Up Assessment

☐ Timing of follow up assessment: specify N/A no follow up assessment

☐ Number of participants included in the follow up assessment: specify N/A no follow up assessment

☐ Consistency of assessment method used: specify N/A no follow up assessment

Rating for **Follow Up Assessment** (select 0, 1, 2, or 3): ☐ 3 ☐ 2 ☐ 1 ☐ <mark>0</mark>
No follow up carried out, outcomes measures at 4 months, 7-9 months and 12 months.

III. Other Descriptive or Supplemental Criteria to Consider
A. External Validity Indicators

A1. Sampling procedures described in detail   <mark>yes</mark> ☐   no ☐
Specify rationale for selection: *the 78 participating schools represented all five provinces in mainland Finland. Schools were stratified by province and language. Participating schools located throughout Finland resembled comprehensive schools throughout the country in such characteristics as class size and proportion of immigrant students.*

Specify rationale for sample size: *Not mentioned*

A1.1 Inclusion/exclusion criteria specified   ☐ <mark>yes</mark> ☐ no

A1.2 Inclusion/exclusion criteria similar to school practice ☐ <mark>yes</mark> ☐ no

A1.3 Specified criteria related to concern ☐ <mark>yes</mark> ☐ no
Universal intervention so included all pupils in a school. Bullying is thought to be a problem in all schools, so all schools would have a goal to reduce bullying.

A2. Participant Characteristics Specified for Treatment and Control Group

Procedural and Coding Manual 64

Participants from Treatment Group	Sample size	Gender	Ethnicity or Birth whole	Ethnic Identity	Race(s)	Acculturation	Primary Language	SES	Family Structure	Locale	Disability	Functional Impairment
☐ Child/Grandson ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other												
☐ Child/Grandson ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other												
☐ Child/Grandson ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other												
☐ Child/Grandson ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other												

Participants from Control Group	Sample size	Gender	Ethnicity or Birth whole	Ethnic Identity	Race(s)	Acculturation	Primary Language	SES	Family Structure	Locale	Disability	Functional Impairment
Child/Grandson Parent/caregiver Teacher												
☐ School ☐ Other ☐ Child/Grandson ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other												
☐ Child/Grandson ☐ Parent/caregiver Teacher												
☐ School ☐ Other ☐ Child/Grandson ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other												

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes ☐ **yes** ☐ no

Specify: age, gender, ethnicity, setting

A3.2 Have relevance to inclusion criteria ☐ yes ☐ no

Specify: *age, gender, ethnicity, setting*

A4. Receptivity/acceptance by target participant population (treatment group)

Participants from Treatment Group	Results (What person reported to have gained from participation in program)	General Rating
☐ Child/Student ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other		☐ Participants reported benefiting overall from the intervention ☐ Participants reported not benefiting overall from the intervention
☐ Child/Student ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other		☐ Participants reported benefiting overall from the intervention ☐ Participants reported not benefiting overall from the intervention
☐ Child/Student ☐ Parent/caregiver ☐ Teacher ☐ School ☐ Other		☐ Participants reported benefiting overall from the intervention ☐ Participants reported not benefiting overall from the intervention

A5. Generalization of Effects:

A5.1 Generalization over time

A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated ☐ yes ☐ no

Specify: *not mentioned*

A5.1.2 Procedures for maintaining outcomes are specified ☐ yes ☐ no

Specify: *not mentioned*

A5.2 Generalization across settings

A5.2.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context: ☐ yes ☐ no

Specify: *whole school intervention: just discussing outcomes in school context. Not home e.g. cyberbullying.*

A5.2.2 Documentation of efforts to ensure application of intervention to other settings ☐ yes ☐ no

Specify: *whole school intervention.*

A5.2.3 Impact on implementers or context is sustained ☐ yes ☐ no

Specify: *No follow up stated.*

A5.3 Generalization across persons

Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated ☐ yes ☐ no

Specify: *Sample group thought to resemble comprehensive schools in Finland.*

B. **Length of Intervention** (select B1 or B2)

B1. ☐ Unknown/insufficient information provided

B2. ☐ Information provided (if information is provided, specify one of the following:)

B2.1 weeks _____
N

B2.2 months _____
N

B2.3 years intervention was 1 school year (however intervention is intended as a long term solution which would be carried out as a whole school approach indefinitely)
N

B2.4 other _____
N

C. **Intensity/dosage of Intervention** (select C1 or C2)

C1. ☐ Unknown/insufficient information provided

C2. ☐ Information provided (if information is provided, specify both of the following:)

C2.1 length of intervention session *20 hour of lessons, other components not specified.*
N

C2.2 frequency of intervention session *10 double lessons throughout the school year.*
N

D. **Dosage Response** (select D1 or D2)

D1. ☐ Unknown/insufficient information provided

D2. ☐ Information provided (if information is provided, answer D2.1)

D2.1 Describe positive outcomes associated with higher dosage: _____

54

E. Program Implementer (select all that apply)

E1. ☐ Research Staff
E2. ☐ School Specialty Staff
E3. ☐ Teachers
E4. ☐ Educational Assistants
E5. ☐ Parents
E6. ☐ College Students
E7. ☐ Peers
E8. ☐ Other
E9. ☐ Unknown/insufficient information provided

F. Characteristics of the Intervener

F1. ☐ Highly similar to target participants on key variables (e.g., race, gender, SES)
F2. ☐ Somewhat similar to target participants on key variables
F3. ☐ Different from target participants on key variables
Not stated

G. Intervention Style or Orientation (select all that apply)

G1. ☐ Behavioral
G2. ☐ Cognitive-behavioral
G3. ☐ Experiential
G4. ☐ Humanistic/interpersonal
G5. ☐ Psychodynamic/insight oriented
G6. ☐ other (specify): Multifaceted theoretical background; studies on social standing of aggressive children and research on the participant role of bullying. Social cognitive theory is used as the framework for understanding the process of social behaviour.
G7. ☐ Unknown/insufficient information provided

H. Cost Analysis Data (select G1 or G2)

H1. ☐ Unknown/insufficient information provided
H2. ☐ Information provided (if information is provided, answer H2.1)

   H2.1 Estimated Cost of Implementation:_____

I. Training and Support Resources (select all that apply)

I1. ☐ Simple orientation given to change agents
I2. ☐ Training workshops conducted

   # of Workshops provided 2

   Average length of training full day

   Who conducted training (select all that apply)

      I2.1 ☐ Project Director
      I2.2 ☐ Graduate/project assistants

55

I2.3 ☐ Other (please specify):
I2.3 ☐ <mark>Unknown</mark>

I3. ☐ <mark>Ongoing technical support</mark> network of school teams were created, consisting of 3 school teams each. The network met 3 times during the school year with one person from the KiVa project guiding the network.
I4. ☐ <mark>Program materials obtained</mark> 20 hours of student lesson plans, symbols (bright vests, posters), presentation graphics, parent guide including information about bullying and advice.
I5. ☐ Special Facilities
I6. ☐ Other (specify):

J. Feasibility

J1.   Level of difficulty in training intervention agents (select one of the following)

J1.1 ☐ High
J1.2 ☐ Moderate
J1.3 ☐ Low
J1.4 ☐ <mark>Unknown</mark>

J2.   Cost to train intervention agents (specify if known): *not known*

J3.   Rating of cost to train intervention agents (select one of the following)

J3.1 ☐ High
J3.2 ☐ Moderate
J3.3 ☐ Low
J3.4 ☐ <mark>Unknown</mark>

Summary of Evidence for Group-Based Design Studies

Indicator	Overall Evidence Rating NNR = No numerical rating or 0 - 3	Description of Evidence Strong Promising Weak No/limited evidence or Descriptive ratings
General Characteristics		
General Design Characteristics	NNR	Volunteering schools in Finland completely randomized experiment.
Statistical Treatment	NNR	Sample size 8232. Sufficiently large N for multilevel modelling.
Type of Program	NNR	Universal programme with indicated actions.
Stage of Program	NNR	First large scale evaluation of the KiVa programme.
Concurrent/Historical Intervention Exposure	NNR	No/ limited evidence.
Key Features		
Measurement	2	Promising evidence. Reliability coefficients above .7 for all but one measure, valid measures referenced in the literature, peer and self-report.
Comparison Group	2	Promising evidence. Randomly assigned control group. Missing data inputted using SAS Proc MI.
~~Primary/Secondary Outcomes are Statistically Significant~~	N/A	
~~Educational/clinical significance~~	N/A	
Identifiable Components	0	No/ limited evidence. Not clear which component of the programme led to decrease in bullying/ victimisation.
Implementation Fidelity	2	Promising evidence. Formal training and KiVa resources given. Network groups for ongoing support.

57

Replication	N/A	N/A
Site of Implementation	N/A	N/A
Follow Up Assessment Conducted	0	No/ limited evidence. No follow up assessment conducted. Data collected at 3 stages (1) 4 months (2) 7-9 months (3) 12 months.
Descriptive or Supplemental Criteria		
External validity indicators	NNR	Carried out in comprehensive schools in Finland as reviewing the effectiveness of KiVa in these target schools.
Length of Intervention	NNR	1 year (although pupils only in school for 9 months of the year due to holidays).
Intensity/dosage	NNR	20 hours (10 double lessons) of student lessons across the year. Other components not stated.
Dosage Response	NNR	Same dose for each school: whole school intervention with outlined components.
Program Implementer	NNR	Teacher to implement student lessons, other components carried out by a range of school staff.
Characteristics of the Intervener	NNR	No/ limited evidence.
Intervention Style/Orientation	NNR	Multifaceted theoretical background; studies on social standing of aggressive children and research on the participant role of bullying. Social cognitive theory is used as the framework for understanding the process of social behaviour.
Cost Analysis Data Provided	NNR	No/ limited evidence.
Training and Support Resources	NNR	2 full days of face-to-face training days. Training networks of school teams created who meet with one person from KiVa project 3 times a year.
Feasibility	NNR	No/ limited evidence.

Appendix E

Details on Criteria and the Rationale for all WoE Ratings

WoE A: Methodological Quality

WoE A is the generic judgement about the quality of the study in line with quality standards for that type of design (Gough, 2007). In this case, group design studies. In order to establish the methodological quality of each study, the amended version of Kratochwill's (2003) APA Task Force coding protocol for group designs (See Appendix B and C) was used to appraise each study. Numerical values for four key dimensions 'measurement', 'comparison group', 'implementation fidelity' and 'follow up assessment' were produced using the criteria outlined in Kratchowill's (2003) coding manual. Each dimension received a numerical value (0-3) depending on how well it met the criteria. A score of 3 indicated 'strong evidence', 2 indicated 'promising evidence, 1 indicated 'weak evidence' and 0 indicated 'no/ limited evidence'. The average score from each section was then calculated to provide an overall measure of methodological quality (WoE A). Table 1 provides the classification of each score achieved in WoE and Table 2 outlines the scores for WoE A.

Table 1

Classification of Score

Classification of Score	Averaged Scores
High	≥ 2.5
Medium	1.4 - 2.5
Low	≤ 1.4

Table 2

Scores for WoE A

Authors	Dimensions				Overall WoE A
	Measurement (0-3)	Comparison Group (0-3)	Implementation Fidelity (0-3)	Follow up Assessment (0-3)	
Kärnä et al., (2011a)	2	2	2	0	1.5 Medium
Nocentini & Mensini (2016)	1.5	2	2	0	1.38 Low
Kärnä et al. (2013)	2	2	2	0	1.5 Medium
Hutchings & Clarkson (2015)	1.5	0	2	0	.88 Low
Kärnä et al., (2011b)	1.5	2	1.5	0	1.25 Low

<u>WoE B: Methodological Relevance</u>

WoE B is a specific judgement about how suitable the research design is for answering the review question (Gough, 2007). In this case, whether the research design is appropriate for evaluating the effectiveness of the KiVa anti-bullying programme in reducing levels of bullying and victimisation in schools. The criteria for WoE B are based on evidence hierarchies (Guyatt et al., 1995). Guyatt et al's. (1995) hierarchy of evidence suggests that the best research designs for addressing questions about intervention effectiveness will usually require methodological designs which use control groups and random allocation. In contrast to this, research designs using cohort studies, case control studies, cross sectional studies or single case designs are noted to be weaker evidence in which to guide the strengths of recommendations for clinical practice. In other words, research designs with minimal threats to internal validity are thought to provide better evidence than research designs with higher threats to internal validity. Criteria for WoE B are provided in Table 3 and scores are outlined in Table 4.

Table 3

Criteria for WoE B

Weighting	Criteria
High (3)	- The study must have an 'active' control group. - Participants must be randomly allocated to condition group. - The study must collect pre and post measurements for primary outcomes. - The study must have a sample size that is adequate for all statistical analysis[a].
Medium (2)	- The study must have a 'no intervention' control group. - Participants must be randomly allocated to condition group or group equivalence must be established through post-hoc analysis. - The study must collect pre and post measurements for primary outcomes. - The study must have a sample size that is adequate for all statistical analysis.
Low (1)	- The study does not include a control group. - The study must collect pre and post measurements for primary outcomes. - The study may have a sample size that is lower than required for statistical analysis.
Zero (0)	- The study does not meet any of the criteria outlined above.

[a] Adequate sample size was calculated from Cohen (1992) based on a medium effect size and alpha level of 0.05.

Table 4

Scores for WoE B

Authors	Overall WoE B
Kärnä et al., (2011a)	2 Medium
Nocentini & Mensini (2016)	2 Medium
Kärnä et al. (2013)	2 Medium
Hutchings & Clarkson (2015)	1 Low
Kärnä et al., (2011b)	2 Medium

<u>WoE C: Topic Relevance</u>

WoE C is a review specific judgement about how suitable the study is for answering the review question (Gough, 2007). In this case, evaluating whether the KiVa anti-bullying programme is an effective intervention to decrease levels of bullying and victimisation in schools. Judgements were made based on the following rationale:

- Topic: The review is evaluating the effectiveness of the KiVa anti-bullying programme, so the study should utilise the original KiVa anti-bullying programme. If an adapted version of KiVa is used then this will not accurately reflect the effectiveness of the programme.

- Sample: Findings can be generalised more widely if the range of demographics in the sample are deemed representative of the intended population.

- Evidence gathering: The research question refers to levels of bullying and victimisation, therefore outcome measures that measure these key constructs should be used. The use of valid and reliable measures gives the reader confidence that they are measuring the intended constructs and that these results are accurate over time. Validation requires demonstration that the measure has evidence and theory to support the interpretation of test scores, and a reliability score of above .85 provides strong evidence, .7 provides promising evidence and above .5 provides weak evidence (Kratochwill, 2003).

Criteria for WoE C are provided in Table 5 and scores are outlined in Table 6.

Table 5

Criteria for WoE C

Weighting	Criteria
High (3)	- The study must use the original version of the KiVa anti-bullying programme (no adaptions; except language). - The study's sample is deemed representative of the intended population and must include at least two of the following demographics (gender, age, ethnicity, socio economic status). - The study must use outcome measures that measure bullying and victimisation. - The study provides evidence that the outcome measures used have a high validity and reliability (r = .85 or higher) for all primary outcome measures.
Medium (2)	- The study must use the KiVa anti-bullying programme. - The study must include at least two of the following demographics (gender, age, ethnicity, socio economic status). - The study must use outcome measures that measure bullying and victimisation. - The study provides evidence that at least 75% of the primary the outcome measures used are valid and have a reliability above r = .7.
Low (1)	- The study uses KiVa anti- bullying programme. - The study must include at least one of the following demographics (gender, age, ethnicity, socio economic status). - The study must use outcome measure that measure bullying and victimisation. - The study provides evidence that the outcome measures used are valid or have a reliability above r = .5.
Zero (0)	- The study does not meet any of the criteria outlined above.

Table 6

Scores for WoE C

Authors	Overall WoE C
Kärnä et al., (2011a)	3 High
Nocentini & Mensini (2016)	2 Medium
Kärnä et al. (2013)	3 High
Hutchings & Clarkson (2015)	2 Medium
Kärnä et al., (2011b)	3 High

WoE D: Overall assessment

WoE D is the combination of Woe A, WoE B and WoE C to form an overall assessment score, which represents the extent to which a study provides evidence to answer the review question (Gough, 2007). This was calculated by averaging the values of WoE A, WoE B and WoE C. Scores for WoE D are outlined in Table 7.

Table 7

Scores for WoE D

Authors	WoE A Methodological Quality	WoE B Methodological Relevance	WoE C Topic Relevance	WoE D Overall weight of evidence
Kärnä et al., (2011a)	1.5 Medium	2 Medium	3 High	2.17 Medium
Nocentini & Mensini (2016)	1.37 Low	2 Medium	2 Medium	1.79 Medium
Kärnä et al. (2013)	1.5 Medium	2 Medium	3 High	2.17 Medium
Hutchings & Clarkson (2015)	.88 Low	1 Low	2 Medium	1.29 Low
Kärnä et al., (2011b)	1.25 Low	2 Medium	3 High	2.08 Medium