

Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

How effective are school-based universal preventive interventions at reducing eating disorder risk factors in adolescent males?

Summary

Eating disorders are a growing concern amongst adolescent populations. It has been argued that school-based universal preventive interventions are a convenient way to try and address this issue due to their wide-reaching access to adolescents in society through the education system. Interventions of this kind can be administered to whole cohorts of adolescents with the aim of ameliorating risk factors of eating disorders, preventing acquisition of eating disorders in later life.

Research around eating disorders typically focuses on females due to larger numbers of diagnoses in this population, but males are becoming more of a focus due to their increasing rates of diagnoses. Research suggests though that preventive interventions for eating disorders are developed around findings focussed on females. This review analysed studies to assess how effective these kinds of interventions are with male adolescents.

After in depth analysis of five studies, it was found that these interventions have little impact on reducing risk factors for eating disorders with male adolescents. Issues regarding the measures used by the studies and bias towards the needs of females were discussed, as well as areas for future development.

Introduction

The prevalence of eating disorders in adolescent populations has been of significance for many years and is still an ongoing area of research (Maganto, Garaigordobil, & Kortabarria, 2016; Muise, Stein, & Arbess, 2003; Neumark-Sztainer, 1996; Shisslak, Crago, Neal, & Swain, 1987).

The UK charity Beating Eating Disorders (BEAT) estimated that in 62% of diagnoses of eating disorders in the UK, symptoms were first exhibited under the age of 16 and that in 24% of diagnoses, symptoms were first exhibited between the ages of 16 and 19 (BEAT, 2015). Adolescent populations are a key risk group for eating disorders (Evans et al., 2017; Maganto et al., 2016; Striegel-Moore & Bulik, 2007) and thus it is ideal to work with these populations in a preventive way.

Although eating disorders are substantially more common with females than with males, estimates place the percentage of males suffering with eating disorders at between 10-25% of the total number of cases in the UK (NHS, 2015; BEAT, 2015); US studies have found higher estimates with males suffering eating disorders placed at between 25% and 36% of estimated cases (Hudson, Hiripi, Pope, & Kessler, 2007). Eating disorders are often observed to carry a stigma with males specifically as these disorders are typically associated with females, which makes figures about the prevalence of these disorders with males unlikely to be accurate as the evidence suggests that male sufferers are less likely to seek help and thus receive a diagnosis (Collier, 2013; Griffiths et al., 2015; Strother, Lemberg, Stanford, & Turberville, 2012).

With eating disorders being treated in clinical settings, much of the research in the area has been based on clinical samples and interventions; research into the role that schools can play in the prevention of eating disorders is substantially less. The efficacy of school-based preventive interventions for eating disorders has had mixed reviews. Schools have been identified as a logical place to implement universal preventive interventions because of the wide population it gives access to (Coie et al., 1993) however, O'Dea (2000) found examples in which school-based preventive interventions increased eating disorder behaviours within their treatment populations. Research regarding the effectiveness of school-based universal preventive interventions is not yet conclusive and requires further exploration.

Universal preventive interventions are used when interventions can benefit all members of a population without having adverse effects (Coie et al., 1993). When considering the psychological role in the prevention of mental disorders, theories of risk and protective factors are crucial. Risk factors in maladaptive thoughts and behaviours mostly involve a number of contributing risk factors, many of which may not be specifically related to the observed disorder (Coie et al., 1993), as they are more general in nature and are causal in a number of maladaptive behaviours.

Universal preventive interventions specific to eating disorder risk factors tend to draw on two main psychological approaches. Many of the interventions take a psychoeducational stance, which have been shown to be effective in providing knowledge on eating disorders, leading to reductions in eating disorder behaviours (Cororve Fingeret, Warren, Cepeda-Benito & Gleaves, 2006). Other interventions work on a dissonance-based approach, used to cause attitudinal change in relation

to eating disorder risk factors noted to specifically target the internalised “thin-ideal” of Western culture, which is thought to reduce risk factors such as body dissatisfaction amongst other eating disorder risk factors (Stice, Shaw, Becker & Rhode, 2008). Many interventions use a combination of both psychoeducational and dissonance-based approaches in the prevention of eating disorder risk factors, and so both approaches are worthy of further investigation in developing the understanding in this area.

Educational Psychologists have the potential to play a crucial role in preventive interventions for eating disorders. As noted, schools are well placed to offer preventive interventions as they give wide access to the child and adolescent populations which are critical to target. Despite this however, school staff alone may not be best equipped to meet these needs. Staff who would often deliver education on health, fitness and nutrition have themselves been found to be more likely at risk of eating disorders (Yager & O’Dea, 2009), and school staff in general lack specialist training in how to manage education about eating disorders (Yager & O’Dea, 2005). Educational Psychologists are highlighted as being crucial in work around eating disorders in schools, notably as providers of information and direction to support (Hellings & Bowles, 2007), therefore working to fill this gap of expertise in schools. As such, educational psychology needs to ensure that this area is one that is sufficiently researched.

Historically, much of the research around eating disorders has been focussed on females (Muise et al., 2003; Strother et al., 2012; Vo, Lau, & Rubinstein, 2016) due to their higher rates of diagnosis. Notably, the prevalence of eating disorders in male

populations is on the increase, although it is possible that this is due to changes in diagnostic criteria and better recognition of symptoms rather than actual increases in eating disorders themselves (Akgul et al., 2016; Hoek & Van Hoeken, 2003). Research is now beginning to attend more to the needs of males suffering with eating disorders (Akgul et al., 2016; Dakanalis et al., 2016; Muise et al., 2003; Vo et al., 2016) and the differences in male and female experiences of eating disorders (Maganto et al., 2016). Evidence suggests that a large number of the risk factors are irrespective of gender such as body dissatisfaction, self-esteem and negative affectivity (Dakanalis et al., 2016; Maganto et al., 2016; Muise et al., 2003) but there are some key differences; self-objectification, particularly around perceived weight and amount of muscle, has been found to be a key risk factor in male adolescents (Dakanalis et al., 2016; Maganto et al., 2016). Actual weight history is also a key factor with males (Akgul et al., 2016; Muise et al., 2003; Strother et al., 2012; Vo et al., 2016) with many exhibiting weight-controlling behaviours to compensate for this, such as dieting behaviours or Binge-Eating behaviours (Dakanalis et al., 2016; Muise et al., 2003).

On reflection of the available evidence, there is an increased need to focus on male adolescent populations at risk of eating disorders. With the rate of diagnosis on the rise within this population, schools are becoming more important in ameliorating the array of potential risk factors adolescents face. The emphasis on the needs of female populations in both the research and in interventions leaves it unclear how effectively the needs of male adolescents are met in this area. This study aims to address this imbalance in the research and treatment by reviewing the effectiveness

of school-based universal preventive interventions for eating disorders at reducing eating disorder risk factors with male adolescents.

Critical Review of the Evidence Base

Systematic Literature Search

A systematic literature search was completed on 20/1/17 using the search-string outlined in Table 1 on the Psycinfo and Medline online databases. As shown in Figure 1, 34 journal articles were identified from this initial search. Based on the inclusion and exclusion criteria outlined in Table 2, 11 of these articles were discounted based on a screening of the titles and one duplicate journal article was identified and removed. After screening the remaining abstracts, 12 articles were removed and reasons for these are identified in Appendix A. After reading through the remaining articles, five were excluded with reasons also identified in Appendix A. Five articles were selected for in-depth analysis in this review and are identified and summarised in Table 3.

Table 1

Search Terms for Obtaining Studies

Search Number	Search Term	Additional Information for Search
1.	eating disorders	Map Term to Subject Heading [Exploded]
2.	"eating disorder" OR bulimia OR anorexia	[title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
3.	1 OR 2	Combine with OR
4.	school based intervention	Map Term to Subject Heading
5.	"school based intervention"	[title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
6.	4 OR 5	Combine with OR
7.	boy* OR male* OR men	[title, abstract, heading word, table of contents, key concepts, original title, tests & measures]
Final Search	3 AND 6 AND 7	Combine with AND

Table 2

Inclusion and Exclusion Criteria for Selecting Relevant Studies

Category	Inclusion Criteria	Exclusion Criteria	Rationale
1. Type of Publication	Article must be from a journal	Article is not from a published journal e.g. publication is a book, unpublished dissertation etc.	Journal articles must meet sufficient academic criteria for reliability
2. Participants	Data for male participants must be present	No specific data for male participants given e.g. data not split by gender or no male participants in study	The research question examines male adolescents exclusively
3. Intervention	Participants must be aged between 11-18 Must be a preventive intervention Must specifically target risk factors of eating disorders	Participants above the age of 18 or below the age of 11 present in study Intervention cannot work with populations with pre-existing diagnosed eating disorders Diagnosable conditions, both physical and mental, which are comorbid to eating disorders must not be treated alongside eating disorder symptoms within the intervention Intervention includes provision taking place outside of a school setting	The research question is looking at school-based preventive interventions for eating disorders exclusively The treatment of comorbid diagnosis may account for some of the variance in eating disorder scores if the variables are closely related, which would make comparisons less accurate.
4. Study Design	Must be school-based Must present empirical primary data	Must not contain secondary data e.g. review papers and meta-analyses	Primary data is required for in-depth analysis
5. Year of Publication	Must have an outcome variable which measures eating disorder risk factors Control groups must be used Papers must have been published no earlier than 2002	No eating disorder risk factors are measured No control groups are used Papers published before 2002	Outcomes will need to show whether eating disorder risk factors have been reduced Required for comparison to measure effectiveness The review should be based on recent data sets due to changes in the research field over time

Figure 1. Article Screening Process

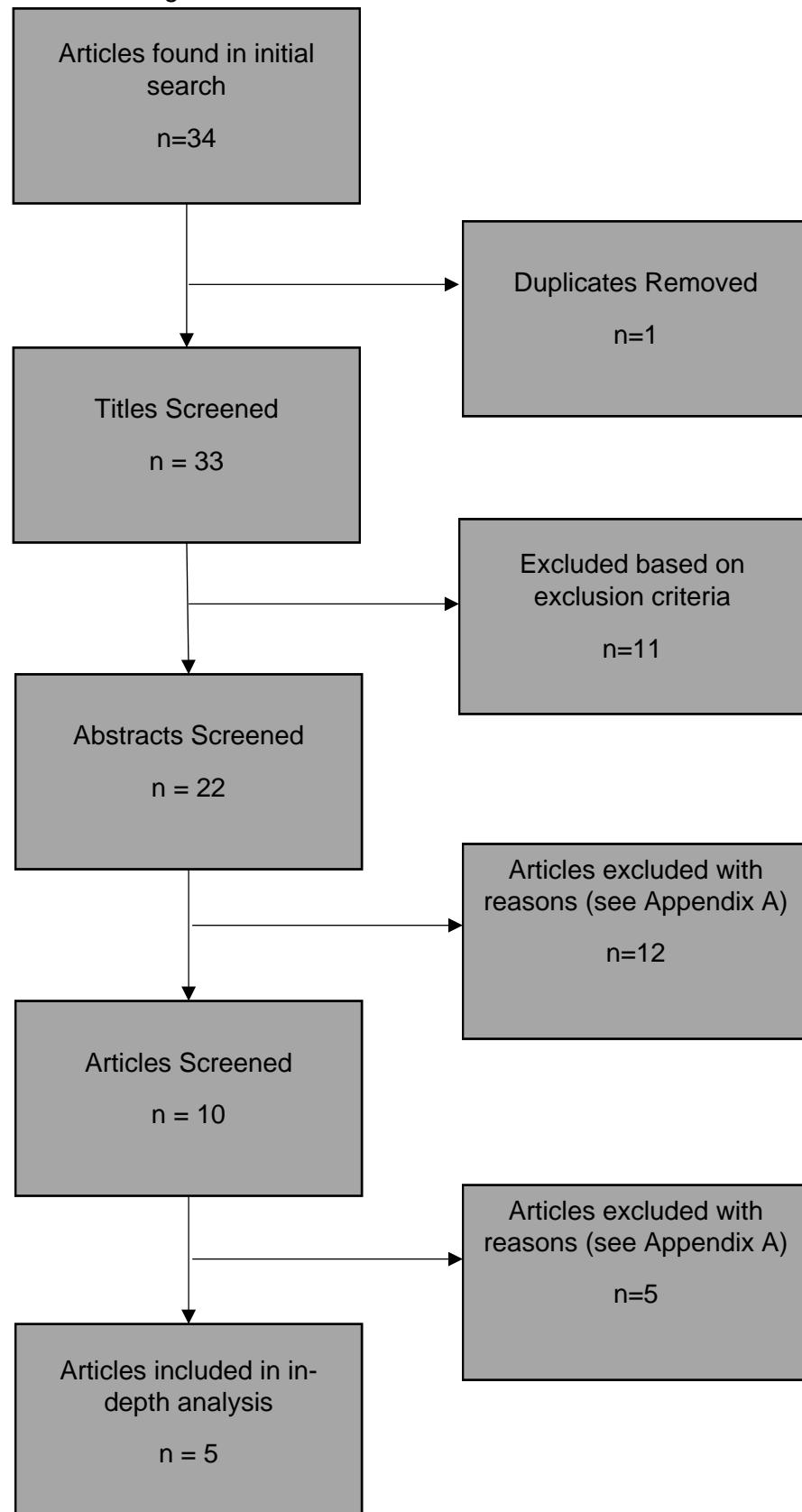


Table 3
Details of Highlighted Studies for In-Depth Analysis

Study	Intervention	Male Sample	Design	Measures	Findings
Berger et al. (2014)	Torera 9 x 90 minute sessions Teacher-led Manualised approach, after initial training	n=256 22 German schools: 7 th grade Mean age = 12.03 years (SD =0.65) ^a 11-13 years old criterion for eligibility	Quasi-experimental design with pre and post measures Schools self-selected into the intervention group (10 schools in total) Two control groups (12 schools in total): CG1 had opted to administer no ED interventions at any point; CG2 had administered different ED interventions prior to Seventh Grade but were not administering any ED preventative interventions in Seventh Grade	Eating Behaviour (SCOFF and the short version of the German version of EAT) Body Self-Esteem (Subscale from the German Body Experience Questionnaire) BMI	Males were not affected by the intervention except in one measure of eating behaviours when comparing baselines to follow up data
			Pre and post intervention measures		
Diedrichs, Atkinson, Steer, Garbett, Rumsey, Halliwell (2015)	Dove Confident Me: Single Session 1 x 90 minute intervention Researcher-led and Teacher-led conditions (Teachers received 2 hour	n=847 6 UK schools: Years 7 and 8 Mean age =12.11 (0.71), 12.3 (0.69), 12.12 (0.71) across each intervention group	3-arm cluster randomised control trial (control; intervention teacher-led; intervention researcher-led) Follow-ups averaged at: control 7 weeks; teacher- led 7.5 weeks; researcher-led 7 weeks	Body esteem (Body Esteem Scale for adolescents and adults) Body satisfaction (Project-EAT III Body Areas Satisfaction Scale) Internalisation of appearance ideals (Sociocultural Attitudes Towards Appearance Questionnaire-3) Sociocultural pressures (Purpose-built measure derived from existing scales of	Dietary restraint was only found to be significantly affected for males when comparing both intervention conditions; no significant

Table 3
Details of Highlighted Studies for In-Depth Analysis

Study	Intervention	Male Sample	Design	Measures	Findings
training session from researchers)				sociocultural pressures) Social comparisons (Social Comparison to Models and Peers Scale) Appearance teasing (Project EAT-III Teasing Scale) Appearance conversations (Appearance conversations with friends subscale of the culture among friends) Negative affect (10-item positive and negative affect schedule for children) Self-esteem (Rosenberg Self-esteem Scale shortened) Dietary restraint (Dutch Eating Behaviour Questionnaire – Restraint) Eating disorder symptoms (SCOFF) Life engagement (Purpose-built measure assessing the extent that worries or feeling bad about the way you look has stopped you, or are likely to stop you, from engaging in life activities)	differences were found between intervention and control conditions

Table 3
Details of Highlighted Studies for In-Depth Analysis

Study	Intervention	Male Sample	Design	Measures	Findings
McVey, Tweed & Blackmore (2007)	Healthy Schools-Healthy Kids 8 month intervention Workshops to school staff and parents Teacher-led curriculum 1 researcher-led focus group session for males only Varied implementation of intervention	n=332 4 matched Canadian schools: 6 th and 7 th grade Mean age = 11.27 years (SD = .67) ^a	Randomised control trial Measures taken at baseline, post-intervention and 6 month follow-up	Body Satisfaction (6 item version of the Body Satisfaction Scale) Internalization of media ideals (Internalization sub-scale of the Sociocultural Attitudes Towards Appearance Questionnaire) Body Size Acceptance (4 item scale from the Sociocultural Attitudes Towards Appearance Questionnaire) Perceptions of appearance or weight-based teasing (The Perception of Teasing Scale: Adolescent Scale)	This intervention had no significant impact on male participants
Pokrajac-Buljan, Živčić-Bećirević, Calugi & Grave (2006)	Eating disorders prevention. An educational and prevention program for social and health operators 6x90 min	n=69 2 Croatian schools: 7 th grade Mean age=12.8 ^a	1 participant school and 1 matched control school Measures taken 1 week before intervention, 1 week after intervention and at 6 months post-intervention	Disordered Eating (ChEAT: Children's Eating Attitude Test) Dieting Behaviour (ADS: Adolescent Dieting Scale) Knowledge of Eating Disorder Risks (The Knowledge Questionnaire)	Disordered eating and dieting behaviour were both significantly improved in male samples of

Table 3
Details of Highlighted Studies for In-Depth Analysis

Study	Intervention	Male Sample	Design	Measures	Findings
Wilksch (2015)	sessions (1 per week)			Self-Esteem (Coopersmith Self-Esteem Scale)	the intervention group
	Psychologist-led			BMI	
	Media-Smart	n=26 (9 in intervention; 15 in control)	Random allocation of control and intervention conditions	Shape and weight concern (Eating Disorder Examination – Questionnaire)	
	8x50 min sessions (2 per week)	2 Australian classes in one school: 7 th grade	Measures taken for baseline, post intervention (5 weeks after baseline) and a follow-up at 6 months	Dietary restraint (Dutch Eating Behaviour Questionnaire – Restraint)	
	Teacher led (training provided)	Mean age = 12.43 years, SD = 0.61 ^a		Body dissatisfaction (EDI: Body dissatisfaction) ^b	
				Media internalization (Sociocultural Attitudes Towards Appearance Questionnaire-3)	
				Feelings of ineffectiveness (EDI: Feelings of ineffectiveness)	
				Depression (Child Depression Inventory-Short Form)	
				Self-esteem (Rosenberg Self-Esteem Scale)	
				Weight-related peer teasing (McKnight Risk Factor Survey)	

^aMean age also included females in these studies – no data was given specifically for the age of the male subgroup

^bBoys received a different scale from female participants on this measure

Weight of Evidence

The highlighted studies identified for in-depth analysis were reviewed for their methodological quality and appropriateness to the research question using Gough's (2007) Weight of Evidence framework.

Weight of Evidence A demonstrates the generic methodological value of the study irrespective of the research question. For the purposes of this review a modified version of Kratochwill's (2003) coding protocols were used to assess this objectively. The protocol was modified to reflect that only male populations would be considered and sections were removed as they were not relevant to the highlighted studies (e.g. sections on qualitative studies). Concepts of primary and secondary outcomes were removed from the coding protocol due to the inconsistencies in the research base in defining clear primary outcomes for measures of eating disorders (see *Measures* section for further discussion).

Weight of Evidence B and C both are specific to the research question of this study and thus have been created for this review; Appendix B includes criteria for both with rationale.

Weight of Evidence B differs from A in that it is specific to the review question, examining the relevance of the designs of the study and the evidence gathered to address the research question.

Weight of Evidence C analyses how appropriate the focus of each included study is to the review question, differing from Weight of Evidence B in that it is more focussed on the thematic aspects of the studies rather than the methodological aspects.

By taking a mean of the scores for Weight of Evidence A-C to calculate Weight of Evidence D, an overall weighting can be applied to each study in order to assess how strong the evidence provided from each of the studies is in an objective and systematic way (Table 4).

Table 4
Weight of Evidence Ratings

Study	Weight of Evidence A	Weight of Evidence B	Weight of Evidence C	Weight of Evidence D
Berger et al. (2014)	Medium (2)	Medium (2.3)	Medium (2)	Medium (2.1)
Diedrichs, et al. (2015)	Low (1)	High (3)	High (3)	Medium (2.3)
McVey et al. (2007)	Low (1.3)	High (2.7)	Medium (2)	Medium (2)
Pokrajac-Bulian, et al. (2006)	Medium (1.7)	High (2.7)	Medium (2.3)	Medium (2.2)
Wilksch (2015)	Medium (1.7)	High (3)	Medium (2.3)	Medium (2.3)

Samples

The samples used within the studies largely varied: Three of the studies used a number of different schools within their intervention and control groups (Berger et al., 2014; Diedrichs et al., 2015; McVey, Tweed, & Blackmore, 2007); Pokrajac-Bulian, Živčić-Bećirević, Calugi, & Dalle Grave (2006) used only 2 schools (one for control and one for intervention); Wilksch, (2015) used only one school, with one class for a control group and one for the intervention.

Sufficient sample sizes were established according to Cohen's (1992) criteria, estimating that, to detect a medium effect size for a power of .80, 64 participants would be needed in each study. Wilksch's was the only study out of the five to not

have a sufficiently large population within the study, with only 26 male participants in total (9 in the intervention group and 15 in the control group).

Intervention and Control Conditions

All of the highlighted studies used clusters when assigning participants to intervention and control conditions rather than complete randomisation; this was due to the school-based nature of the studies. In their use, these interventions would be administered to pre-existing classes and so replicating this in the studies rather than wholly random assignment gives an accurate representation of how effective these interventions could be outside of an experimental design.

Randomisation is considered best practice to minimise statistical bias (Flay et al., 2005). The majority of the five highlighted studies used randomisation when assigning participant clusters to the control or intervention conditions (Diedrichs et al., 2015; McVey et al., 2007; Wilksch, 2015) although two did not (Berger et al., 2014; Pokrajac-Bulian et al., 2006).

All of the control conditions used the same approach of administering no parallel or delayed intervention, they all completed submitted data without any change in their day-to-day schooling which is considered adequate to demonstrate an effect of the intervention (Flay et al., 2005).

Interventions were administered by a range of people. Teachers led the Torera, Healthy Schools-Healthy Kids and Media Smart interventions (Berger et al., 2014; McVey et al., 2007; Wilksch, 2015) and researchers led Eating disorders prevention:

An educational and prevention program for social and health operators (Pokrajac-Bulian et al., 2006). Dove Confident Me (Diedrichs et al., 2015) had two separate control conditions: one teacher-led and the other researcher-led.

Measures

All measures used within the highlighted studies were self-reported by participants, including measures of BMI. The lack of triangulation of the data means that we should treat outcomes with caution as measuring outcomes from a range of sources provides data which are more reliable and valid (Flay et al., 2005).

Across all five studies, 27 different outcomes were measured using a wide variety of standardised and unstandardised instruments. For the purpose of this review, three areas of outcomes were identified for further analysis, taken from the Eating Disorder Inventory (Garner, 2004) instrument. This instrument is comprised of twelve subscales, three of which are classed as Eating Disorder Specific scales: Drive for Thinness, Bulimia and Body Dissatisfaction. Using Garner's (2004) descriptors of these three subscales 14 of the original 27 outcomes were mapped on to three areas, demonstrated in Table 5, and these outcomes were used for the purpose of this study; the remaining outcomes will not be addressed within this review.

By focussing on these eating disorder-specific outcomes, this study aims to address the research question more closely by excluding confounding variables which, although may be classed as a risk factor for eating disorder by the studies, are not exclusively linked with eating disorders and could be related to a number of other unrelated diagnoses or difficulties (Coie et al., 1993).

Table 5
Measures mapped to EDI-3 Eating Disorder Specific Skills

Eating Disorder Specific Scale from Eating Disorder Inventory-3	Outcomes from selected studies mapped to scale descriptors from Garner (2004)
Drive for Thinness	Eating Behaviour Eating Disorder Symptoms Dietary Restraint Body Size Acceptance Disordered Eating Weight-loss Dieting Behaviour Shape and weight concern
Bulimia	Eating Behaviour Eating Disorder Symptoms Dietary Restraint Disordered Eating Dieting Behaviour
Body Dissatisfaction	Body Esteem Body Self-Esteem Body Satisfaction Body Dissatisfaction Body Size Acceptance Muscle Gain Shape and weight concern

Fidelity

In order to gain an accurate view of how effective each intervention is, it is key to consider how each one was implemented within the studies, as deviations away from standardised implementation could potentially confound measures of its success.

Only one of the highlighted studies demonstrated high fidelity according to the fidelity criterion in Weight of Evidence C (see Appendix B); Dove: Confident Me (Diedrichs et al., 2015) had a rigid manualised approach and teacher and researcher-led conditions were reported separately; fidelity measures were reported within the study and showed good adherence to the source material during the intervention.

Two studies demonstrated moderate levels of fidelity according to the fidelity criterion of Weight of Evidence C; Healthy Schools-Healthy Kids (McVey et al., 2007) used a manualised approach for delivery and all teachers were trained in its use.

The study included fidelity checks but no discussion of the findings of these were mentioned within the paper other than one note saying “teachers were not compliant in recording which activities they conducted with students.” Eating disorders prevention: An educational and prevention program for social and health operators (Pokrajac-Bulian et al., 2006) also used a manualised approach and reported on the effectiveness of its translation from Italian to Croatian for the purposes of the study. The study noted that it was two of the researchers who administered the intervention and that both received the necessary training prior to this but no further attempts to ensure fidelity were discussed.

The remaining two studies failed to demonstrate adequate levels of fidelity in the implementations of their interventions, both receiving a low rating for the fidelity criterion in Weight of Evidence C; Wilksch (2015) identified the need to improve the fidelity of future testing of the Media Smart intervention by using audio-recordings from teacher delivered sessions. Berger et al. (2014) noted that, although the Torera program followed a manualised approach, eight out of ten teachers were considered to “not follow the manual strictly,” which was found to moderate some of the outcomes of the study.

Gender-based characteristics of interventions

With the differences in experiences between male and female experiences of eating disorders, one intervention had gender-specific components to manage the varying needs of each population.

Healthy School-Healthy Kids (McVey et al., 2007) included a sub-intervention specifically aimed at female adolescents called Girl Talk , which consisted of 12 50-minute sessions. Males too received a specifically designed session for them but it was only one 50-minute focus group and focussed on teasing behaviours related to their bodies. The study also only measured disordered eating outcomes for its female participants without providing a rationale for why.

Outcomes

The five studies varied in their inclusion of effect sizes, shown in Table 6; Wilksch (2015) was the only study to provide effect sizes for all outcomes; Berger et al. (2014) gave effect sizes only when a significant effect was measured; Diedrichs et al. (2015) gave effect sizes for specific genders only if a significant relationship between gender and variable was found, otherwise an effect size was calculated for the male and female participants combined Both McVey et al. (2007) and Pokrajac-Bulian et al. (2006) provided means and standard deviations for measured outcomes but did not calculate effect sizes; effect sizes for these studies were calculated using Morris' (2008) Pretest-Posttest-Control Group Effect Size calculation in order to best fit the design of these studies.

As all studies received the same Weight of Evidence D, all effect sizes should be interpreted as being equally useful in answering the research question. Effect sizes were classified as being low, medium or high according to Cohen's (1992) ratings. The majority of the highlighted studies showed no significant change for male pupils as a result of the interventions with the exception of 3 studies, 2 of which demonstrated significant effects for only one measure (Berger et al., 2014; Diedrichs

et al., 2015), and 1 which had significant outcomes for all highlighted measures (Pokrajac-Bulian et al., 2006).

Berger et al. (2014) found a small effect on eating behaviours in male participants at 6 months after the intervention. It is notable though that no significant effect was measured immediately post-intervention, and neither was any significant effect measured for eating behaviours measured using parallel testing of the construct with the SCOFF questionnaire. Although this study had the highest rating of all the highlighted studies for Weight of Evidence A, meaning it was the most methodologically sound, it received the lowest ratings for Weight of Evidence B and C, which suggests that it is less effective in addressing the review question; these results should therefore be interpreted cautiously.

Diedrichs et al. (2015) found a small significant effect for the dietary restraint outcomes of its male participants at both immediately after the intervention and at the follow-up measure. These effects however were only observed when comparing the teacher-led and researcher-led conditions; no significant effects were found between the intervention and control conditions. Of note is that this study received the lowest Weight of Evidence A rating of all of the highlighted studies, suggesting it was not methodologically sound and so the significant results should be interpreted carefully.

Pokrajac-Bulian et al. (2006) found significant effects for all of its highlighted outcomes. The intervention was shown to have a strong effect on disordered eating behaviours, significantly lowering the scores in the male population immediately after

the intervention; at the follow-up the effect was reduced (with an effect size just below the threshold for medium classification) but still significant. The intervention also had a small, significant effect on dieting behaviour, reducing it between the baseline and immediately after the intervention and at the follow-up.

All other studies (McVey et al., 2007; Wilksch, 2015) failed to produce any significant effect on highlighted outcomes as a result of their intervention on the male population within their studies, and any effect sizes calculated for these non-significant results suggested a small effect.

Table 6
Effect Sizes for Highlighted Outcomes

Study	Number of Participants	Variable measure	Effect size location	Time of measure	Significance at 0.05 level	Effect size	Interpretation of effect size ^a	WoE D Rating
Berger et al. (2014)	n=256	Eating Behaviour SCOFF	None reported	t_1	p=ns	-	-	Medium
			None reported	t_2	p=ns	-	-	
		EAT	None reported Extracted from article	t_1 t_2	p=ns p=0.039	t_2 $d=0.35$	Small	
		Body Self-Esteem	None reported None reported	t_1 t_2	p=ns p=ns	-	-	
Diedrichs, Atkinson, Steer, Garbett, Rumsey, Halliwell (2015)	n=847	Body Esteem	None reported	Post/Follow up	p=ns	-	-	Medium
		Body Satisfaction	None reported	Post/Follow up	p=ns	-	-	
		Dietary Restraint	Extracted from article	(researcher-led vs teacher-led) Post Intervention Follow-up	p=0.034 p=0.045	$d=-0.20$ $d=-0.18$	Small Small	
		Eating Disorder Symptoms (SCOFF)	None reported	Post/Follow up	p=ns	-	-	
McVey, Tweed & Blackmore (2007)	n=332	Body Satisfaction	Calculated	Post-intervention Follow-up	p=ns p=ns	$d_{ppc2}=-0.03$ $d_{ppc2}=0.01$	Small Small	Medium
		Body Size Acceptance	Calculated	Post-intervention Follow-up	p=ns p=ns	$d_{ppc2}=0.18$ $d_{ppc2}=0.18$	Small Small	
		Weight-	None reported	Post-intervention Follow-up	p=ns	-	-	

Table 6
Effect Sizes for Highlighted Outcomes

Study	Number of Participants	Variable measure	Effect size location	Time of measure	Significance at 0.05 level	Effect size	Interpretation of effect size ^a	WoE D Rating
loss/muscle gain								Post/Follow-up
Pokrajac-Bulian, Živčić-Bećirević, Calugi & Grave (2006)	n=69	Disordered Eating	Calculated	Post-intervention	p=<0.01	$d_{ppc2}=-0.90$	Large	Medium
		Dieting Behaviour	Calculated	Follow-up	p=<0.05	$d_{ppc2}=-0.46$	Small	
Wilksch (2015)	n=26	Shape and Weight Concern	Extracted from article	Post-intervention	p=<0.01	$d=0.02$	Small	Medium
		Dietary Restraint	-	Follow-up	p=ns	$d=0.04$	Small	
		Body Dissatisfaction	Extracted from article	-	p=ns	$d=0.02$	Small	
			Extracted from article					

^a Small .2; Medium .5; Large .8 (Cohen, 1992)

Conclusions and Recommendations

This review aimed to assess the effectiveness of school-based universal preventive interventions for eating disorders at reducing eating disorder risk factors with male adolescents. Based on the findings of the highlighted studies, and in particular the large proportion of non-significant results and low effect sizes, there seems to be little support for how these interventions work with this population. Across all of the studies, the most common finding was that the interventions had very little impact if any across the majority of the highlighted outcomes with only 30% of them being significant and only one of the highlighted measures having an effect size classified as greater than small. Further to this, all of the highlighted studies produced a medium Weight of Evidence D rating, meaning that none of the studies provided strong evidence in relation to research question which adds strength to the argument that this area of research requires further investigation. Possible reasons for these findings are discussed further in this section, as are the further implications for future research and work within this area.

One notable point raised in Pokrajac-Bulian et al.'s (2006) study was the rationale for including males in the study. Although this study in review was the most effective at reducing eating disorder risk factors with its male population, males were included to teach them "what they can do to help a friend or another person affected by these disorders" and to "play an important role in reducing the social pressure on women to be thin and beautiful." Despite this rationale, there was no evidence towards fostering these ideals within the intervention itself as all participants regardless of gender received identical inputs. Furthermore, McVey et al.'s (2007) study only monitored disordered eating in females and offered males one individual 50 minute

session compared to the 12 50-minute sessions for the females. This supports findings that research focuses on female populations, which offers a potential explanation for the lack of measured impact (Collier, 2013; Griffiths et al., 2015; Strother et al., 2012).

The measures used across these studies may also explain the lack of impact found. Pokrajac-Bulian et al. (2006) and Berger et al.'s (2014) studies were the only to report on disordered eating in males, both finding some significant effects. Notably research shows that this risk factor is key within adolescent males as eating behaviours become problematic when self-objectification is an issue (Dakanalis et al., 2016), which may explain why it was shown to have been significantly affected. The exclusion of this measure in other studies may explain why significant effects were not shown. Pokrajac-Bulian et al. (2006) were also the only researchers to also measure dieting behaviours rather than dieting restraint. The same evidence regarding male disordered eating could be applied to dieting behaviour. Dieting restraint is not considered a key risk factor for males as they tend to exhibit more binge-eating behaviours (Dakanalis et al., 2016), which would most likely be reflected in dieting behaviour measures.

On reflection of these findings, it is possible that it is the measures taken within studies such as these which dictates how effective they appear to be.

The wide-range of possible measures for eating disorder risk factors is a potential confounding issue in the study of this area. Within the five highlighted studies, 27 individual risk factors were identified. Given that all of the interventions aim for the

same outcome, the lowering of eating disorder risk factors, it suggests that there is a lack of research into which eating disorder risk factors should be prioritised to be a successful intervention. For males, dieting behaviours have been highlighted as a more significant risk factor (Dakanalis et al., 2016; Striegel-Moore & Bulik, 2007), which this review provides some evidence towards; future research should examine how effective preventive interventions for eating disorders are at ameliorating this risk factor with male adolescent populations.

The plethora of potential risk factors to an eating disorder themselves is suggestive that universal preventive interventions as a concept may not be appropriate to support with this need. Interventions may need to be targeted specifically to the people considered most at risk.

In conclusion, school-based universal preventive interventions show very little evidence to suggest that they are able to reduce eating disorder risk factors with adolescent males. In order to improve the standards of outcomes for this group, more research needs to focus on this population in terms of priority risk factors and effective measures of these to inform future interventions tailored specifically to their needs.

References

- Akgul, S., Akdemir, D., Kara, M., Derman, O., Cetin, F. C., & Kanbar, N. (2016). The understanding of risk factors for eating disorders in male adolescents. *International Journal of Adolescent Medicine and Health*, 28(1), 97–105. <https://doi.org/10.1515/ijamh-2014-0078>
- BEAT. (2015). The costs of eating disorders, Social, health and economic impacts, (February), 76. Retrieved from https://www.b-eat.co.uk/assets/000/000/302/The_costs_of_eating_disorders_Final_original.pdf?1424694814
- Berger, U., Schaefer, J. M., Wick, K., Brix, C., Bormann, B., Sowa, M., ... Strauss, B. (2014). Effectiveness of Reducing the Risk of Eating-Related Problems Using the German School-Based Intervention Program, "Torera", for Preadolescent Boys and Girls. *Prevention Science*, 15(4), 557–569. <https://doi.org/10.1007/s11121-013-0396-4>
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112(1), 155–159.
- Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., ... Long, B. (1993). The Science of Prevention. *American Psychologist*, 48(10), 1013–1022. <https://doi.org/10.1037/0003-066X.48.10.1013>
- Collier, R. (2013). Treatment challenges for men with eating disorders. *Canadian Medical Association Journal*, 185(3), 137–139. <https://doi.org/10.1503/cmaj109-4363>
- Cororve Fingeret, M., Warren, C. S., Cepeda-Benito, A., & Gleaves, D. H. (2006). Eating disorder prevention research: a meta-analysis. *Eating Disorders*, 14(3), 191–213.
- Dakanalis, A., Clerici, M., Caslini, M., Gaudio, S., Serino, S., Riva, G., & Carr, G. (2016). Predictors of initiation and persistence of recurrent binge eating and inappropriate weight compensatory behaviors in college men. *International Journal of Eating Disorders*, 49(6), 581–590. <https://doi.org/10.1002/eat.22535>
- Diedrichs, P. C., Atkinson, M. J., Steer, R. J., Garbett, K. M., Rumsey, N., & Halliwell, E. (2015). Effectiveness of a brief school-based body image intervention "Dove Confident Me: Single Session" when delivered by teachers and researchers: Results from a cluster randomised controlled trial. *Behaviour Research and Therapy*, 74, 94–104. <https://doi.org/10.1016/j.brat.2015.09.004>
- Evans, E. H., Adamson, A. J., Basterfield, L., Le Couteur, A., Reilly, J. K., Reilly, J. J., & Parkinson, K. N. (2017). Risk factors for eating disorder symptoms at 12 years of age: A 6-year longitudinal cohort study. *Appetite*, 108, 12–20. <https://doi.org/10.1016/j.appet.2016.09.005>
- Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6(3), 151–175. <https://doi.org/10.1007/s11121-005-5553-y>
- Garner, D. M. (2004). Eating Disorder Inventory-3 (EDI-3). *International Journal of Eating Disorders*, 35, 478–479.

- 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. <https://doi.org/10.1080/02671520701296189>
- Griffiths, S., Mond, J. M., Li, Z., Gunatilake, S., Murray, S. B., Sheffield, J., & Touyz, S. (2015). Self-stigma of seeking treatment and being male predict an increased likelihood of having an undiagnosed eating disorder. *International Journal of Eating Disorders*, 48(6), 775–778. <https://doi.org/10.1002/eat.22413>
- Hellings, B., & Bowles, T. (2007). Understanding and Managing Eating Disorders in the School Setting. *Australian Journal of Guidance and Counselling*, 17(01), 60–67.
- Hoek, H. W., & Van Hoeken, D. (2003). Review of the Prevalence and Incidence of Eating Disorders. *International Journal of Eating Disorders*, 34(4), 383–396. <https://doi.org/10.1002/eat.10222>
- Hudson, J. I., Hiripi, E., Pope, H. G., & Kessler, R. C. (2007). The Prevalence and Correlates of Eating Disorders in the National Comorbidity Survey Replication. *Biological Psychiatry*, 61(3), 348–358. <https://doi.org/10.1016/j.biopsych.2006.03.040>
- Kratochwill, T. R. (2003). Task Force on Evidence-Based Interventions in School Psychology, (608). Retrieved from http://www.indiana.edu/~ebi/documents/_workingfiles/EBImanual1.pdf
- Maganto, C., Garaigordobil, M., & Kortabarria, L. (2016). Eating Problems in Adolescents and Youths: Explanatory Variables. *The Spanish Journal of Psychology*, 19(e81), 1–9. <https://doi.org/10.1017/sjp.2016.74>
- McVey, G., Tweed, S., & Blackmore, E. (2007). Healthy Schools-Healthy Kids: A controlled evaluation of a comprehensive universal eating disorder prevention program. *Body Image*, 4(2), 115–136. <https://doi.org/10.1016/j.bodyim.2007.01.004>
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational research methods*, 11(2), 364–386.
- Muisse, A. M., Stein, D. G., & Arbess, G. (2003). Eating disorders in adolescent boys: A review of the adolescent and young adult literature. *Journal of Adolescent Health*, 33(6), 427–435. [https://doi.org/10.1016/S1054-139X\(03\)00060-0](https://doi.org/10.1016/S1054-139X(03)00060-0)
- Neumark-Sztainer, D. (1996). School-based programs for preventing eating disturbances. *The Journal of School Health*, 66(2), 64–71. <https://doi.org/10.1111/j.1746-1561.1996.tb07912.x>
- NHS (2015, June 29) Eating Disorders. Retrieved from <http://www.nhs.uk/conditions/eating-disorders/pages/introduction.aspx>
- O'Dea, J. A. (2000). School-based interventions to prevent eating problems: First do no harm. *Eating Disorders*, 8(2), 123–130. <https://doi.org/10.1080/10640260008251219>
- Pokrajac-Bulian, A., Živčić-Bećirević, I., Calugi, S., & Dalle Grave, R. (2006). School prevention program for eating disorders in Croatia: A controlled study with six

- months of follow-up. *Eating and Weight Disorders*, 11(4), 171–178.
<https://doi.org/http://dx.doi.org/10.1007/BF03327568>
- Shisslak, C. M., Crago, M., Neal, M. E., & Swain, B. (1987). Primary prevention of eating disorders. *Journal of Consulting and Clinical Psychology*, 55(5), 660–7.
<https://doi.org/10.1017/CBO9780511543890.023>
- Stice, E., Shaw, H., Becker, C. B., & Rohde, P. (2008). Dissonance-based interventions for the prevention of eating disorders: Using persuasion principles to promote health. *Prevention Science*, 9(2), 114.
- Striegel-Moore, R. H., & Bulik, C. M. (2007). Risk factors for eating disorders. *American Psychologist*, 62(3), 181–198. <https://doi.org/10.1111/j.1749-6632.1997.tb48199.x>
- Strother, E., Lemberg, R., Stanford, S. C., & Turberville, D. (2012). Eating disorders in men: Underdiagnosed, undertreated, and misunderstood. *Eating Disorders*, 20(5), 346–355. <https://doi.org/10.1080/10640266.2012.715512>
- Vo, M., Lau, J., & Rubinstein, M. (2016). Eating Disorders in Adolescent and Young Adult Males: Presenting Characteristics. *Journal of Adolescent Health*, 59(4), 397–400. <https://doi.org/10.1016/j.jadohealth.2016.04.005>
- Wilksch, S. M. (2015). School-based eating disorder prevention: A pilot effectiveness Trial of Teacher-delivered Media Smart. *Early Intervention in Psychiatry*, 9(1), 21–28. <https://doi.org/10.1111/eip.12070>
- Yager, Z., & O'Dea, J. A. (2005). The role of teachers and other educators in the prevention of eating disorders and child obesity: What are the issues? *Eating Disorders*, 13(3), 261–278. <https://doi.org/10.1080/10640260590932878>
- Yager, Z., & O'Dea, J. (2009). Body image, dieting and disordered eating and activity practices among teacher trainees: Implications for school-based health education and obesity prevention programs. *Health Education Research*, 24(3), 472–482. <https://doi.org/10.1093/her/cyn044>

Appendix A Studies excluded according to inclusion criteria

Study	Inclusion criteria violated
Yager, Z., Diedrichs, P. C. and Drummond, M. (2013). Understanding the role of gender in body image research settings: Participant gender preferences for researchers and co-participants in interviews, focus groups and interventions. <i>Body Image, 10</i> (4) 574-582.	3. Intervention (No intervention) 4. Study Design (Incorrect measures)
Mennuti, R. B., Bloomgarden, A., Mathison, J. & Gabriel, N. (2006) Adolescents with eating disorders: School-based cognitive-behavioral interventions. In Mennuti, R. B., Freeman, A., & Christner, R. W. (Eds.) <i>Cognitive-behavioral interventions in educational settings: A handbook for practice.</i> (pp275-303). Psychology Press.	1. Type of Publication
O'Dea, J. A., & Yager, Z. (2011). School-based psychoeducational approaches to prevention.	1. Type of Publication
Yager, Z. & O'Dea, J. A. (2008). Prevention programs for body image and eating disorders on university campuses: A review of large, controlled interventions. <i>Health Promotion International, 23</i>(2), 173-189.	4. Study Design (Review Paper)
Tate, A. (2007) Schooling. In Lask, B., & Bryant-Waugh, R. (Eds.). <i>Eating disorders in childhood and adolescence 3rd edition.</i> (pp312-334). Routledge.	1. Type of Publication
Myers, L. L. (2006). Anorexia Nervosa, Bulimia Nervosa, and Obesity. <i>Fostering Child and Adolescent Mental Health in the Classroom, 141.</i>	1. Type of Publication
O'Dea, J. A. (2005). School-based health education strategies for the improvement of body image and prevention of eating problems: An overview of safe and successful interventions. <i>Health Education, 105</i>(1), 11-33.	4. Study Design (Review Paper)
Wade, T. D., Davidson, S., & O'Dea, J. A. (2002). Enjoyment and perceived value of two school-based interventions designed to reduce risk factors for eating disorders in adolescents. <i>Australian e-Journal for the advancement of mental health, 1</i>(2), 113-120.	4. Study Design (Incorrect measures)
Varnado-Sullivan, P. J., Zucker, N., Williamson, D. A., Reas, D., Thaw, J., & Netemeyer, S. B. (2001). Development and implementation of the Body Logic Program for adolescents: A two-stage prevention program for eating disorders. <i>Cognitive</i>	3. Intervention (Non-school-based component)

and Behavioral Practice, 8(3), 248-259.

- Jones, M., Lynch, K. T., Kass, A. E., Burrows, A., Williams, J., Wilfley, D. E., & Taylor, C. B. (2014). Healthy weight regulation and eating disorder prevention in high school students: a universal and targeted Web-based intervention. *Journal of medical Internet research, 16*(2), e57.
- Klein, E. G., Lytle, L. A., & Chen, V. (2008). Social ecological predictors of the transition to overweight in youth: results from the Teens Eating for Energy and Nutrition at Schools (TEENS) study. *Journal of the American Dietetic Association, 108*(7), 1163-1169.
- Varnado-Sullivan, P. J., & Zucker, N. (2004). The Body Logic Program for adolescents: A treatment manual for the prevention of eating disorders. *Behavior modification, 28*(6), 854-875.
- D'Souza, C. M., Forman, S. F., & Austin, S. B. (2005). Follow-up evaluation of a high school eating disorders screening program: knowledge, awareness and self-referral. *Journal of Adolescent Health, 36*(3), 208-213.
- Haines, J., Neumark-Sztainer, D., & Morris, B. (2008). Theater as a behavior change strategy: Qualitative findings from a school-based intervention. *Eating disorders, 16*(3), 241-254.
- Martinsen, M., Bahr, R., Børresen, R., Holme, I., Pensgaard, A. M., & Sundgot-Borgen, J. (2014). Preventing eating disorders among young elite athletes: a randomized controlled trial. *Medicine & Science in Sports & Exercise, 46*(3), 435-447.
- Wilksch, S. M., Tiggemann, M., & Wade, T. D. (2006). Impact of interactive school-based media literacy lessons for reducing internalization of media ideals in young adolescent girls and boys. *International Journal of Eating Disorders, 39*(5), 385-393.
- Wilksch, S. M., & Wade, T. D. (2014). Depression as a moderator of benefit from Media Smart: A school-based eating disorder prevention program. *Behaviour research and therapy, 52*, 64-71.

3. Intervention (No school-based component)

3. Intervention (does not target eating disorder risk factors)
4. Study Design (Incorrect measures)

3. Intervention (No intervention)

4. Study Design (No control group)

2. Participants (no separate data for males; wrong age)
4. Study Design (No control group)

3. Intervention (uses participants already diagnosed with eating disorders)

4. Study Design (No control group)

2. Participants (no separate data for males)

Appendix B Materials for generating Weight of Evidence B and C Ratings

WoE B

1. Randomised assignment of participant clusters to control or intervention group required for valid and reliable findings
2. Outcomes must be measured at baseline, at post-intervention and a follow-up measure should be taken to measure impact of outcomes over time
3. Outcomes must include standardised measures of eating disorder risk factors for valid and reliable findings

Descriptor	Criteria
High (3)	<ol style="list-style-type: none"> 1. Participant clusters are randomly assigned to intervention and control conditions 2. Outcomes must be measured pre and post intervention, and a follow up measure must be taken 3. All relevant outcomes are standardised measures
Medium (2)	<ol style="list-style-type: none"> 1. Participant clusters are not randomly assigned to intervention and control conditions 2. Outcomes must be measured pre and post intervention 3. Relevant outcomes include standardised and non-standardised measures
Low (1)	<ol style="list-style-type: none"> 1. No control condition is present for assignment 2. Effectiveness is judged on observations/measures made post-study only 3. All relevant outcomes are non-standardised measures

Study	Criteria Scoring	Final WoE B score ^a (mean criteria score)
Berger et al. (2014)	<ol style="list-style-type: none"> 1. Medium (2) 2. Medium (2) 3. High (3) 	Medium (2.3)
Diedrichs et al. (2015)	<ol style="list-style-type: none"> 1. High (3) 2. High (3) 3. High (3) 	High (3)
McVey et al. (2007)	<ol style="list-style-type: none"> 1. High (3) 2. High (3) 3. Medium (2) 	High (2.7)
Pokrajac-Bulian, et al. (2006)	<ol style="list-style-type: none"> 1. Medium (2) 2. High (3) 3. High (3) 	High (2.7)
Wilksch (2015)	<ol style="list-style-type: none"> 1. High (3) 2. High (3) 3. High (3) 	High (3)

^aDescriptors: Low 0-1.4; Medium 1.5-2.4; High 2.5-3

WoE C

1. Clusters must implement the intervention in the same way to ensure fidelity; Implementation should adhere strictly to intervention materials/guidance
2. Interventions should be implemented the same for participants of any gender so that no bias towards gender impacts on findings
3. All three eating disorder-specific scales from the Eating Disorder Inventory-3 have at least one corresponding eating disorder risk factor measured in the study in order to best review specific eating disorder outcomes

Descriptor	Criteria
High (3)	<ol style="list-style-type: none"> 1. High fidelity stated in the study 2. All participants in the intervention condition receive identical intervention 3. Eating disorder risk factors measures map across all three eating disorder-specific scales from the Eating Disorder Inventory
Medium (2)	<ol style="list-style-type: none"> 1. High fidelity implied OR Medium fidelity stated or implied 2. Participants within the intervention condition receive some differentiation of intervention according to gender 3. Eating disorder risk factors measures map across two of the eating disorder-specific scales from the Eating Disorder Inventory
Low (1)	<ol style="list-style-type: none"> 1. Low fidelity implied or stated 2. The intervention condition offers a completely different intervention for each gender 3. Eating disorder risk factors measures map across only one of the eating disorder-specific scales from the Eating Disorder Inventory

Study	Criteria Scoring	Final WoE C score^a (mean criteria score)
Berger et al. (2014)	<ol style="list-style-type: none"> 1. Low (1) 2. Medium (2) 3. High (3) 	Medium (2)
Diedrichs, Atkinson, Steer, Garbett, Rumsey, Halliwell (2015)	<ol style="list-style-type: none"> 1. High (3) 2. High (3) 3. High (3) 	High (3)
McVey, Tweed & Blackmore (2007)	<ol style="list-style-type: none"> 1. Medium (2) 2. Medium (2) 3. Medium (2) 	Medium (2)
Pokrajac-Bulian, Živčić-Bećirević, Calugi & Grave Wilksch	<ol style="list-style-type: none"> 1. Medium (2) 2. High (3) 3. Medium (2) 1. Low (1) 2. High (3) 3. High (3) 	Medium (2.3)

^aDescriptors: Low 0-1.4; Medium 1.5-2.4; High 2.5-3

Appendix C Materials for generating Weight of Evidence A

The coding protocol used below is an adapted version of Kratochwill's (2003) used to assess the general strength of the study design. The coding protocol was adapted by extracting the three key areas of measurement, comparison group and appropriate statistical analysis rather than the entire coding protocol. These areas were selected for the following reasons:

- Measurement: As there are many different outcomes measured as risk factors of eating disorders, it is important to assess the strength of the outcomes measured in each of the studies. In doing so in this review, it allows for more effective comparisons to be drawn between all of the disparate measures used by interpreting the more valid and reliable measures as more significant.
- Comparison Group: The need for a control group was highlighted in the inclusion criteria as being necessary to ensure that the measured effect of an intervention could be clearly attributed to the intervention itself. In doing so, all of the studies in this review will have comparison groups but some may be used more effectively than others, which could potentially impact on their findings, therefore it is important to assess how effectively comparison groups have been used when interpreting each study's findings.
- Appropriate Statistical Analysis: Consistency between how researchers run statistical analyses on their data is not guaranteed. By selecting this criterion in the coding protocol, it helps to ensure that the data itself is more comparable as it will give more weight to studies who have used more appropriate statistical analyses.

Weight of evidence A was calculated by giving these three key areas in each study a rating of 0-3 (3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence). The mean of these three scores were then used as a final Weight of Evidence A score. The criteria for assigning these scores, taken from Kratochwill's (2003) paper, are given below.

Score for Measure	Criteria
Strong Evidence (3)	If multiple measures are used, they must all meet these criteria. <ol style="list-style-type: none"> 1. Measures must produce reliable scores (at least .85) for the majority of primary outcomes and the type of reliability statistic used must be explicit. 2. Data must be collected from multiple sources. 3. Data must be collected via multiple methods. 4. Researchers demonstrate that the measures used assess the primary outcomes.
Promising Evidence (2)	If multiple measures are used, 75% must meet these criteria. <ol style="list-style-type: none"> 1. Measures must produce reliable scores (at least .70). 2. Data must be collected from multiple sources and/or via multiple methods.
Weak Evidence (1)	If multiple measures are used, 50% must meet these criteria. <ol style="list-style-type: none"> 1. Measures must produce reliable scores (at least .50).
No evidence (0)	1. Reliability scores are <.50

Score for Comparison Group	Criteria
Strong Evidence (3)	<ol style="list-style-type: none"> 1. Uses a minimum of one “active” comparison group. 2. Group equivalence must be established, preferably through random assignment to intervention and comparison groups. 3. Change agents must be counterbalanced. 4. Equivalent mortality and low attrition at post-intervention (and at follow-up where appropriate)
Promising Evidence (2)	<ol style="list-style-type: none"> 1. Uses a minimum of one “no intervention” comparison group. 2. Two of the following must be met: group equivalence is established; change agents are counterbalanced; equivalent mortality with low attrition. 3. If equivalent mortality is not established then no significant group differences between control and intervention groups must be found after an intent-to-intervene analysis.
Weak Evidence (1)	<ol style="list-style-type: none"> 1. Uses a comparison group. 2. One of the following must be met: counterbalancing of change agents; group equivalence is established; equivalent mortality with low attrition.
No evidence (0)	<ol style="list-style-type: none"> 1. No attempt to establish group equivalence.

Score for Appropriate Statistical Test	Criteria
Strong Evidence (3)	<ol style="list-style-type: none"> 1. Appropriate unit of analysis 2. Family-wise error rate controlled 3. Sufficiently large population
Promising Evidence (2)	<ol style="list-style-type: none"> 1. Two of the following: Appropriate unit of analysis; family-wise error rate controlled; sufficiently large population.
Weak Evidence (1)	<ol style="list-style-type: none"> 1. One of the following: Appropriate unit of analysis; family-wise error rate controlled; sufficiently large population.
No evidence (0)	<ol style="list-style-type: none"> 1. None of the following: Appropriate unit of analysis; family-wise error rate controlled; sufficiently large population.

Appendix D: Example of a Coding Protocol calculating Weight of Evidence A

Coding Protocol

Full Study Reference in proper format: Berger, U., Schaefer, J. M., Wick, K., Brix, C., Bormann, B., Sowa, M., Strauss, B. (2014). Effectiveness of Reducing the Risk of Eating-Related Problems Using the German School-Based Intervention Program, "Torera", for Preadolescent Boys and Girls.

Prevention Science, 15(4), 557–569. <https://doi.org/10.1007/s11121-013-0396-4>

Intervention Name (description of study): Torera

- Type of Publication:
- Book/Monograph
- Journal Article
- Book Chapter
- Other (specify):

1. General Characteristics

A. General Design Characteristics

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

- Completely randomized design
- Randomized block design (between participants, e.g., matched classrooms)
- Randomized block design (within participants)
- Randomized hierarchical design (nested treatments)
- Randomized cluster design

A2. Nonrandomized designs (if non-random assignment design, select one of the following)

- Nonrandomized design
- Nonrandomized block design (between participants)
- Nonrandomized block design (within participants)
- Nonrandomized hierarchical design
- Nonrandomized cluster design
- Quasi-experimental designs

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

- Very low (little basis)
- Low (guess)
- Moderate (weak inference)
- High (strong inference)
- Very high (explicitly stated)
- N/A
- Unknown/unable to code

B. Participants

Total size of sample (start of study): 256

Intervention group sample size: 97

Control group sample size: CG1=105 CG2=54

C. Type of Program

- Universal prevention program
- Selective prevention program
- Targeted prevention program
- Intervention/Treatment
- Unknown

D. Stage of Program

- Model/demonstration programs
- Early stage programs
- Established/institutionalized programs
- Unknown

E. Concurrent or Historical Intervention Exposure

- Current exposure
- Prior exposure
- Unknown

2. Key Features for Coding Studies and Rating Level of Evidence/Support

(Rating Scale: 3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence)

A. Measurement (Estimating the quality of the measures used to establish effects)

A1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes

- Yes
- No
- Unknown/unable to code

A2 Multi-method (at least two assessment methods used)

- Yes
- No
- N/A
- Unknown/unable to code

A3 Multi-source (at least two sources used self-reports, teachers etc.)

- Yes
- No
- N/A
- Unknown/unable to code

A4 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)

- Yes validated with specific target group

- In part, validated for general population only
- No
- Unknown/unable to code

Overall Rating for measurement: 2

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence

B. Comparison Group

B1 Type of Comparison Group (Select one of the following)

- Typical intervention (typical intervention for that setting, without additions that make up the intervention being evaluated)
- Attention placebo
- Intervention element placebo
- Alternative intervention
- Pharmacotherapy
- No intervention
- Wait list/delayed intervention
- Minimal contact
- Unable to identify type of comparison

B2 Overall confidence of judgment on type of comparison group

- Very low (little basis)
- Low (guess)
- Moderate (weak inference)
- High (strong inference)
- Very high (explicitly stated)
- Unable to identify comparison group

B3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc were counter-balanced across intervention)

- By change agent
- Statistical (analyse includes a test for intervention)
- Other
- Not reported/None

B4 Group equivalence established (select one of the following)

- Random assignment
- Posthoc matched set
- Statistical matching
- Post hoc test for group equivalence

B5 Equivalent mortality

- Low attrition (less than 20 % for post)
- Low attrition (less than 30% for follow-up)
- Intent to intervene analysis carried out?

Overall rating for Comparison group 1

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence

C. Appropriate Statistical Analysis**Analysis 1 ANCOVA**

- Appropriate unit of analysis
- Familywise/experimenter wise error rate controlled when applicable
- Sufficiently large N

Analysis 2 Sobel test

- Appropriate unit of analysis
- Familywise/experimenter wise error rate controlled when applicable
- Sufficiently large N

Analysis Chi Squared

- Appropriate unit of analysis
- Familywise/experimenter wise error rate controlled when applicable
- Sufficiently large N

Overall rating for appropriate statistical analyses 3

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence

Summary of Evidence

Indicator	Overall evidence rating 0-3
Measurement	2
Comparison group	1
Appropriate Statistical Analysis	3