

Case study 1: An Evidence-based practice review report.

Theme: School/Setting Based Interventions for Social, Emotional and Mental Health.

How effective is the body image intervention 'Happy Being Me' in improving body satisfaction for pre-adolescent and adolescent girls?

Summary

Body dissatisfaction among pre-adolescent and adolescent females can undermine academic achievement and hinder opportunities for success in society (Halliwell et al., 2014). This review will evaluate the effectiveness of 'Happy Being Me' (HBM), a universal, school-based, preventative intervention for improving body satisfaction in pre-adolescent and adolescent girls (Richardson & Paxton, 2010). This is particularly relevant at a time when educational psychologists (EPs) are increasingly involved in recommending interventions for Social, Emotional and Mental Health (SEMH) (Hicks et al., 2014). A systematic literature search identified six studies which met the inclusion criteria for the review. The studies were critically appraised using Gough's (2007) Weight of Evidence (WoE) Framework. Following an in-depth evaluation of findings, it is concluded that the evidence for HBM is variable, and effectiveness may be influenced by study design and intervention implementation. The strengths and limitations of the studies are evaluated throughout, and recommendations for EP practice and future research are considered.

Introduction

Body dissatisfaction is considered a public health concern in the United Kingdom (UK) (Government Equalities Office, 2013) and is associated with negative outcomes in young people, including disordered eating, poor self-esteem, and depression (Stewart et al., 2020). Although body dissatisfaction has increased in adolescent males, research suggests adolescent females report higher levels of body dissatisfaction (Dunstan et al., 2017). Body dissatisfaction often manifests in adolescence, however, concerns can emerge earlier, in the pre-adolescent phase (Stewart et al., 2020). For example, UK research found that 48% of girls aged 9-12 were dissatisfied with their bodies (Sands & Wardle, 2003). This highlights a need for early intervention to prevent the onset of body dissatisfaction and reduce negative outcomes (Bird et al., 2013).

Psychological Theory

Evidence suggests that internalisation of the 'thin ideal', body comparison, appearance conversations, and appearance teasing are risk factors for body dissatisfaction (Richardson & Paxton, 2010). Internalisation of the 'thin ideal' involves accepting the media's portrayal of the ideal body and making comparisons with one's own body. A perceived discrepancy between the 'actual' and ideal body can result in body dissatisfaction (Durkin et al., 2007). Adolescents also engage in body comparison in their social environment (Richardson & Paxton, 2010). Jones (2001) found that peers were frequently the targets of body comparisons, and students who reported greater social comparison with peers had higher levels of body dissatisfaction. Additionally, conversations about appearance are frequent and often consist of 'fat talk', which reinforces the importance of appearance, and has been identified as a causal risk factor in body dissatisfaction (Richardson & Paxton, 2010). Finally,

Jones et al. (2004) found a direct association between peer appearance criticism and body dissatisfaction in adolescent girls, suggesting appearance teasing is a risk factor for body dissatisfaction.

Happy Being Me (HBM)

HBM is a preventative, school-based, body image intervention, developed by Richardson and Paxton (2010). HBM targets psychological risk factors for body dissatisfaction including internalisation of the ‘thin ideal’, body comparisons, self-esteem, and environmental risk factors including appearance conversations and weight related teasing (Stewart et al., 2020). HBM was originally intended for adolescent girls, who were considered a high-risk group, and was delivered as three, 50-minute sessions. The content and aims of each session are outlined in Table 1.

Table 1

Aims, Content and Processes Involved in HBM (Summarised from Richardson and Paxton, 2010)

Session	Aims	Content	Processes
1	To increase media literacy. To reduce internalisation of the thin ideal.	The media’s manipulation of images. Appearance and value are not inherently linked. Ideal body differs between cultures and across time.	Combination of worksheets, group discussion, brainstorming, reading and homework.
2	To educate on appearance conversations. To highlight the impact of ‘fat talk’ and appearance	Exploration of the impact of fat talk and appearance teasing and of the strategies	Combination of didactic presentation, role play, brainstorming, and group worksheets.

Session	Aims	Content	Processes
	teasing and develop skills for these situations. To reduce fat talk and appearance teasing.	that could be used in these situations.	
3	To educate on body comparison and increase skills in identifying body comparison. To highlight the consequences of body comparison and develop skills in avoiding body comparison. To revisit media's manipulation of images. To highlight positive qualities not associated with appearance. To review the programme.	Introduction to body comparison and the negative consequences. Explore strategies to use instead of body comparison. Media's manipulation of images. Emphasis on positive qualities not related to appearance. Review sessions.	Combination of presentation, worksheets, group activity, class discussion, film clip presentation, and paired work.

The intervention has since been adapted, for example by increasing the number of sessions, adapting delivery for boys and girls (Dunstan et al., 2017), and adding content related to eating concerns (Wilksch et al., 2015).

Rationale and Relevance to Educational Psychology

A review by Halliwell et al. (2014) concluded that body dissatisfaction has a detrimental impact on girls' confidence and consequently, their academic achievement. A key consideration is the impact of body dissatisfaction on participation in school. A UK study found that 14% of girls aged 11-17, with low body esteem, avoided attending school due to being worried about their looks (The Future Foundation, 2012, as cited in Halliwell et al.,

2014). International studies also show that girls who perceived themselves to be overweight, regardless of actual weight, had lower academic performance than girls who did not (Florin et al., 2011; Mikkilä et al., 2003). The long-term impact of body dissatisfaction has also been investigated. A large-scale study highlighted that 17% of women have avoided job interviews and 8% have avoided going to work due to feeling dissatisfied with their looks (Dove, 2006, as cited in Halliwell et al., 2014). Consequently, there is a clear need for early intervention to address body dissatisfaction and reduce its negative impact on the educational outcomes and aspirations of girls.

The green paper 'Transforming Children and Young People's Mental Health Provision' (Department of Health and Social Care and Department of Education, 2017) promotes the development of a whole-school approach to mental health. This includes identifying needs, selecting and implementing appropriate interventions, and monitoring their effectiveness. The legislation suggests schools could be supported to implement group-based interventions for high-risk adolescent girls, which encourage pupils to critique the 'thin-ideal'.

EPs are in the unique position to explore young people's SEMH, their academic progress, and the relationship between them (Kratochwill, 2007). Given EPs' extensive and relevant training, they are often at the forefront of supporting schools to select, implement, and monitor evidence-based interventions for SEMH (Hicks et al., 2014). If EPs are expected to recommend and support the implementation of evidence-based body image interventions, research should systematically evaluate the effectiveness of programmes and the quality of existing evidence. Therefore, this review poses the following question:

'How effective is the body image intervention 'Happy Being Me' in improving body satisfaction for pre-adolescent and adolescent girls?'

Critical Review of the Evidence Base

Literature Search

A systematic literature search was carried out in November 2020 to identify articles relevant for the review. The search terms outlined in Table 2 were utilised to search across three online databases: PsychInfo, ERIC and the Web of Science.

Table 2

Search Terms Used for Systematic Literature Search

Search Term	Rationale
("happy being me" OR "happy being me intervention" OR "happy being me program*")	This review focuses on a specific, named body-image intervention which is designed for implementation in schools.
("body image" OR "body concern*" OR "body attitude*" OR "body anxiet*" OR "body checking" OR "body esteem" OR "body evaluation" OR "body dissatisfaction" OR "body image disturbance" OR "body surveillance")	This review focuses on an intervention for body image. Articles sometimes do not specify the name of the intervention in the abstract and title and refer instead to a 'body image intervention'.
("child*", OR "young person" OR "youngster" OR "adolescent*" OR "school-age" OR "pre-adolescent" OR "teen*" OR "youth")	This review focuses on the effectiveness of the intervention for pre-adolescent and adolescent girls.

Article Screening

The database searches generated 29 results. Following the removal of duplicates (N=3) and the addition of one article through an ancestral search and one through a scoping search, 28 articles were identified for title and abstract screening. Based on the inclusion and exclusion criteria outlined in Table 3, 19 records were excluded. The remaining nine studies were identified for full-text screening and three records were excluded. The rationale for excluding

these studies, based on inclusion and exclusion criteria, is outlined in Appendix A. The six studies identified as appropriate for the review are outlined in Table 4. The process used to select the final studies is depicted by Figure 1.

Figure 1

A Flow Diagram Outlining the Selection Process

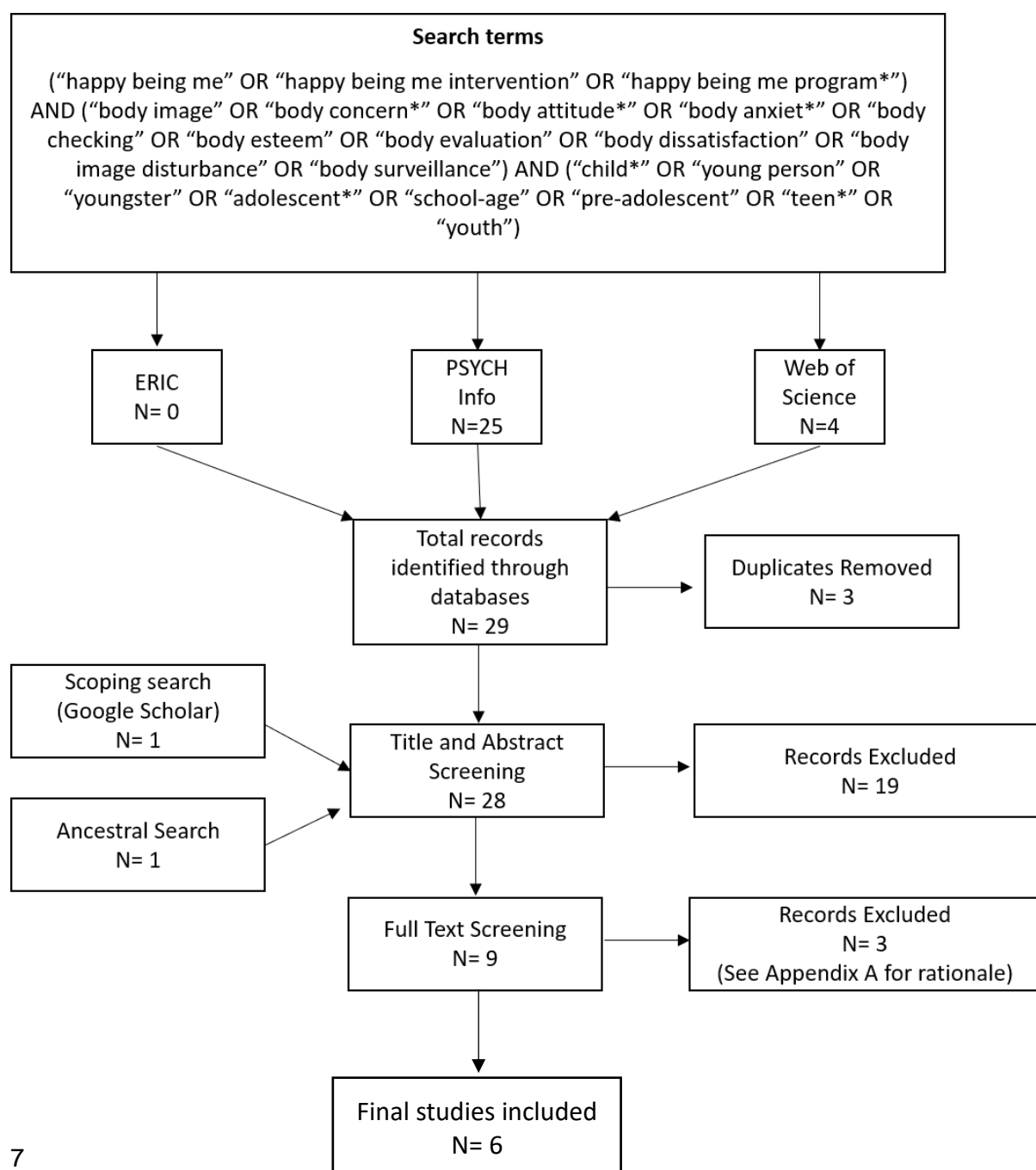


Table 3

Inclusion and Exclusion Criteria for Literature Search

Criteria	Inclusion	Exclusion	Rationale
1. Intervention	Original or adapted version of 'Happy being me' intervention	Any intervention which is not an original or adapted version of 'Happy being me'	This review focuses on an evaluation of the effectiveness of this specific body-image intervention
2. Participants	Aged between 10-18 years	Aged under 10 years or over 18 years	The intervention is focused on pre-adolescent/adolescent age group as this is a high-risk group
3. Participants	Examines effectiveness of intervention for female participants	Only examines effectiveness of intervention for male participants (study does not include females)	Females have been found in the literature to be at highest risk of body dissatisfaction and therefore, the effectiveness for females is the focus of the review
4. Setting	School-based	Any setting other than school for	This intervention is designed to be implemented in schools

Criteria	Inclusion	Exclusion	Rationale
		example, clinic or home.	
5.	Study design Quantitative studies	Qualitative studies	This will enable a review of the effectiveness of the intervention on outcomes in terms of effect sizes
6.	Study design Contains primary empirical data	Study does not contain primary empirical data	The review examines primary empirical data
7.	Study design Experimental design	Non-experimental design	To enable comparisons between groups and the generation of effect sizes
8.	Study design Has a no-intervention or active control group	Studies with no control group or studies which do not include a control which receives non-active components of intervention	This enables examination of effectiveness of the intervention specifically, as opposed to other variables
9.	Outcome measure Include an outcome measure of body dissatisfaction or body satisfaction	Does not include an outcome measure of body dissatisfaction or body satisfaction	This review investigates effectiveness of intervention in reducing body dissatisfaction or improving body satisfaction

Table 4

References of Studies Included in Review

Reviewed Studies	
1.	Richardson, S. M., & Paxton, S. J. (2010). An evaluation of a body image intervention based on risk factors for body dissatisfaction: A controlled study with adolescent girls. <i>International Journal of Eating Disorders</i> , 43(2), 112–122. https://doi.org/10.1002/eat.20682
2.	Bird, E. L., Halliwell, E., Diedrichs, P. C., & Harcourt, D. (2013). Happy being me in the UK: A controlled evaluation of a school-based body image intervention with pre-adolescent children. <i>Body Image</i> , 10(3), 326–334. https://doi.org/10.1016/j.bodyim.2013.02.008
3.	Stewart, C., Goddard, E., Cakir, Z., Hall, R., & Allen, G. (2020). Can more people be “Happy Being Me”? Testing the delivery of a universal body satisfaction program by clinicians and school staff. <i>Eating Disorders</i> , 1–28. https://doi.org/10.1080/10640266.2020.1771165
4.	Dunstan, C. J., Paxton, S. J., & McLean, S. A. (2017). An evaluation of a body image intervention in adolescent girls delivered in single-sex versus co-educational classroom settings. <i>Eating Behaviors</i> , 25, 23–31. https://doi.org/10.1016/j.eatbeh.2016.03.016
5.	Wilksch, S. M., Paxton, S. J., Byrne, S. M., Austin, S. B., McLean, S. A., Thompson, K. M., Dorairaj, K., & Wade, T. D. (2015). Prevention Across the Spectrum: A randomized controlled trial of three programs to reduce risk factors for both eating disorders and obesity. <i>Psychological Medicine</i> , 45(9), 1811–1823. https://doi.org/10.1017/S003329171400289X
6.	McLean, S. A., Wertheim, E. H., Marques, M. D., & Paxton, S. J. (2019). Dismantling prevention: Comparison of outcomes following media literacy and appearance comparison modules in a randomised controlled trial. <i>Journal of Health Psychology</i> , 24(6), 761–776. https://doi.org/10.1177/1359105316678668

Mapping the Field

All studies used group experimental designs to evaluate the effectiveness of HBM. Details of the reviewed studies are provided in Appendix B.

Weight of Evidence (WoE)

Gough’s (2007) WoE framework was used to appraise studies across three dimensions. WoE A considers the methodological quality of the studies. Gersten et al.’s (2005) coding protocol was adapted and used to assess WoE A. Coding protocols designed by the author were used to assess WoE B, which considers methodological relevance to the review question, and WoE C, which considers how appropriate the study topic is for answering the review question. The overall quality and relevance of each study is given by an average across the three dimensions (WoE D). Table 5 summarises WoE ratings for the studies reviewed. Further details related to all WoE criteria rationales is given in Appendix C. The adapted WoE A protocol can be found in Appendix D, and a completed WoE A coding protocol for one study can be found in Appendix E.

Table 5

WoE D Ratings for Reviewed Studies

Study	WoE A	WoE B	WoE C	WoE D (1 Decimal Place)
Richardson & Paxton (2010)	3 (High)	1.5 (Medium)	1.8 (Medium)	2.1 (Medium)
Bird et al. (2013)	3 (High)	1.5 (Medium)	2 (Medium)	2.2 (Medium)
Stewart et al. (2020)	2 (Medium)	1.75 (Medium)	2.8 (High)	2.2 (Medium)
Dunstan et al. (2017)	3 (High)	2.25 (Medium)	2.2 (Medium)	2.5 (High)

Study	WoE A	WoE B	WoE C	WoE D (1 Decimal Place)
Wilksch et al. (2015)	2 (Medium)	2.5 (High)	2.2 (Medium)	2.2 (Medium)
McClellan et al. (2016)	3 (High)	2.75 (High)	1.8 (Medium)	2.5 (High)

Note. WoE D ratings are described as ‘High’ for scores ≥ 2.5 , ‘Medium’ for scores ≥ 1.5 and < 2.5 , and ‘Low’ for scores < 1.5 .

Participants

2427 participants were included in the reviewed studies. Four studies had statistical power to detect small effect sizes, resulting in high WoE B ratings (Dunstan et al., 2017; McLean et al., 2019; Stewart et al., 2020; Wilksch et al., 2015). Two studies were powered to detect medium and large effect sizes, so received medium WoE B ratings (Bird et al., 2013; Richardson & Paxton, 2010). Of the total sample, 1868 were female. The number of females included varied from 42-840. In accordance with the inclusion criteria, the age of participants ranged from 10-14 years.

Four studies sampled participants from more than two schools (Dunstan et al., 2017; McLean et al., 2019; Stewart et al., 2020; Wilksch et al., 2015). These studies received higher WoE C ratings due to the increased generalisability of results, compared to studies who sampled participants from only two schools (Bird et al., 2013; Richardson & Paxton, 2010).

Two studies sampled from single-sex settings (McLean et al., 2019; Richardson & Paxton, 2010). As the majority of UK schools are co-educational (Department for Education (DfE), 2021), findings from single-sex studies may be less generalisable, so these studies received low WoE C ratings for ‘setting generalisability’. One study sampled from a co-educational setting and received a medium rating (Bird et al., 2013). Three studies sampled from a combination of single-sex and co-educational settings (Dunstan et al., 2017; Stewart et al.,

2020; Wilksch et al., 2015) and Stewart et al. (2020) compared outcomes across both settings. These three studies therefore received high WoE C ratings for 'setting generalisability'.

None of the reviewed studies randomly assigned individual participants. Although it can cause classroom disruption, individual random allocation is considered the most effective way of reducing selection bias (Barker et al., 2016). Consequently, none of the reviewed studies received high WoE B ratings for 'participant allocation'. Three studies used block random allocation by class, or school, and received medium ratings. Two of the three studies used teachers to allocate classes to the intervention or control condition (Bird et al., 2013; Stewart et al., 2020), and one study allocated on a first come, first served basis (Richardson & Paxton, 2010). As these three studies did not randomly allocate participants at the individual or block level, they received low WoE B ratings for this criterion.

Location

Two studies were conducted in the UK and received high WoE C ratings for 'location', as findings are most likely to be generalisable to English educational settings (Bird et al., 2013; Stewart et al., 2020). Four studies were conducted in Australia (Dunstan et al., 2017; McLean et al., 2019; Richardson & Paxton, 2010; Wilksch et al., 2015). These studies received medium WoE C ratings, as Australia is economically similar to the UK, and is likely to have a comparable education system. All studies delivered HBM within schools, as required by the inclusion criteria.

Research Design

All studies used a group experimental design. Randomised Controlled Trials (RCTs) are deemed the most appropriate research design for investigating effectiveness questions (Petticrew & Roberts, 2003). Three studies employed an RCT design and received high WoE B ratings for 'design' (Dunstan et al., 2017; McLean et al., 2019; Wilksch et al., 2015). Three studies utilised quasi-experimental designs (Bird et al., 2013; Richardson & Paxton,

2010; Stewart et al., 2020), which are subject to selection bias (Barker et al., 2016). These studies therefore received medium WoE B ratings.

McLean et al. (2019) utilised an active control group (HBM-Eating). This study received the highest WoE B rating for 'control condition', as utilising an active control enabled the researchers to isolate the effects of key intervention components. One study used alternative body image interventions alongside a no-intervention comparison group (Wilksch et al., 2015) and received a medium WoE B rating for 'control condition', as this enabled researchers to control for the effects of factors like attention received from the facilitator, on post-intervention outcomes (Boot et al., 2013). Four studies utilised no-intervention control conditions, resulting in low WoE B ratings for 'control condition' (Bird et al., 2013; Dunstan et al., 2017; Richardson & Paxton, 2010; Stewart et al., 2020).

All six studies used follow-up measurements to evaluate whether there was a sustained effect of HBM, which contributed to WoE A ratings.

Intervention Implementation

Only Richardson and Paxton (2010) implemented HBM in its original format (three, 50-minute sessions). This study received the highest WoE C rating for 'implementation', as it specifically answers the review question about HBM as it was originally designed. The remaining five studies adapted HBM, resulting in medium WoE C ratings for 'implementation'. Details of adaptations can be found in Table 6. All studies clearly described the intervention and delineated any adaptations made, contributing positively to WoE A ratings.

Table 6

Adaptations Made to HBM Intervention

Study	Number of Sessions	Length	Content Adaptations
Bird et al. (2013)	3	60 minutes	Adapted for delivery with males and females.
Dunstan et al. (2017)	6	1 Lesson	Adapted for delivery with 10-11-year-olds. Adapted for delivery in co-educational environment with males and females (content on muscular ideal as well as thin ideal).
Stewart et al. (2020)	6	1 Lesson	Adapted for delivery in co-educational environment with males and females (content on muscular ideal as well as thin ideal).
Wilksch et al. (2015)	8	1 Lesson	Addition of content related to eating concerns and renamed the Helping, Encouraging, Listening and Protecting Peers Initiative (HELPP).
McLean et al. (2019)	3	1 Lesson	Dismantled original format into two separate interventions: HBM-Media and HBM-Comparison.

The fidelity of implementation was only assessed by McLean et al. (2019). Facilitators completed self-report checklists, and rated the proportion of activities delivered and their effectiveness as high. Despite the subjectivity and inherent bias of self-report measures, inclusion of a fidelity assessment contributed positively to WoE A ratings. The remaining five studies did not assess fidelity and received WoE A penalties (Bird et al., 2013; Dunstan et al., 2017; Richardson & Paxton, 2010; Stewart et al., 2020; Wilksch et al., 2015).

In two studies, the researchers delivered HBM (Bird et al., 2013; Richardson & Paxton, 2010) and in three studies, HBM was delivered by trained psychology assistants or students (Dunstan et al., 2017; McLean et al., 2019; Wilksch et al., 2015). These five studies received low WoE C ratings for 'intervention delivery', as delivery by non-school staff is unlikely to reflect how HBM would be implemented in practice. Stewart et al. (2020) compared the delivery of HBM by clinicians and school staff. This study received the highest WoE C rating for 'intervention delivery', as they could evaluate the impact of facilitator on the effectiveness of HBM.

Outcome Measures

All studies included standardised measures of body satisfaction, or body dissatisfaction, that have been found to produce reliable and valid scores with adolescent girls (Durkin et al., 2007; Rodgers et al., 2014). Four studies used the 9-item Eating Disorder Inventory- Body Dissatisfaction Subscale (EDI-BD; Garner et al., 1983) which is rated on a 6-point Likert scale and assesses beliefs about body parts (Dunstan et al., 2017; McLean et al., 2019; Richardson & Paxton, 2010; Stewart et al., 2020; Wilksch et al., 2015). Three studies used the Extended Body Satisfaction Visual Analogue Scale (BSVA; Durkin & Paxton, 2002), which includes five items, rated by participants using a Likert scale (Bird et al., 2013), or by marking responses on a 100-millimetre line (Richardson & Paxton, 2010; Stewart et al.,

2020). Only Richardson and Paxton's (2010) study used both the EDI-BD and the BSVA, contributing positively to WoE A ratings.

Findings

Study findings are summarised in Table 8. Table 7 outlines the effect size descriptors used to interpret values of partial eta squared (η_p^2), based on recommendations by Cohen (1988). A summary of the descriptive statistics for the studies can be found in Appendix F. With the exception of one study (Wilksch et al., 2015), effect sizes were taken from the results sections of articles.

Table 7

Effect Size Descriptors for η_p^2

η_p^2	Descriptor
.14	Large
.06	Medium
.01	Small

Richardson and Paxton (2010) found a medium, significant effect of HBM in reducing girls' body dissatisfaction at post-intervention ($\eta_p^2=.090$, $p<.001$) and 3-month follow-up ($\eta_p^2=.068$, $p<.01$). They also reported a medium, significant effect of HBM in improving body satisfaction at post-intervention ($\eta_p^2=.113$, $p<.001$) and 3-month follow-up ($\eta_p^2=.113$, $p<.001$). This study received a medium WoE D rating, and was the only study to implement HBM in its original format, contributing positively to WoE C ratings. However, its WoE B rating was negatively impacted by the use of a quasi-experimental design and a no-intervention control condition. These limitations should be considered when interpreting findings.

Bird et al. (2013) found a large, significant effect of HBM in reducing body dissatisfaction in girls at post-intervention ($\eta_p^2=.30$, $p<.001$) and 3-month follow-up ($\eta_p^2=.16$, $p<.05$). This study received a medium WoE D rating, and was conducted in the UK, which contributed positively to WoE C ratings. However, this study also utilised a quasi-experimental design and a no-

intervention control condition. Similar caution should therefore be taken when interpreting Bird et al.'s (2013) findings.

Stewart et al. (2020) found a small, significant effect of HBM in improving girls' body satisfaction post-intervention ($\eta_p^2=.01$, $p<.05$). Effect sizes at follow-up were not provided separately by gender, however, the effect of HBM was no longer significant at 3-month follow-up ($\eta_p^2=.02$, $p>.05$). This study received a medium WoE D rating, however, also used a quasi-experimental design and a no-intervention control, resulting in a medium WoE B rating. The study did receive high WoE C ratings on all criteria, contributed to by the comparison of clinician versus school staff delivery of HBM. They found a small, significant, time and group interaction effect on body satisfaction ($\eta_p^2=.012$, $p<.05$), and post-hoc tests demonstrated a significant, medium sized effect of HBM on body satisfaction in the clinician-led group post-intervention ($\eta_p^2=.10$, $p<.001$). The same effect was not observed in the teacher-led group.

Dunstan et al. (2017) found a small, significant effect of HBM in reducing girls' body dissatisfaction ($\eta_p^2=.05$, $p<.05$), which was no longer significant at 6-month follow-up. This study received a high WoE D rating, and the RCT design contributed positively to WoE B ratings. Dunstan et al. (2017) also compared the effectiveness of HBM when delivered in co-educational versus single-sex settings, contributing positively to WoE C ratings. They found no significant effect of group (co-education versus single-sex) on the impact of HBM ($\eta_p^2=.01$, $p>.05$). The high methodological quality and relevance of this study should be considered when interpreting findings.

Wilksch et al. (2015) did not provide sufficient information to extract or calculate an effect size for the impact of HELPP, compared to the control. The study only provided effect sizes for significant interactions and did not include the sample size of intervention and control groups, resulting in WoE A penalties. Wilksch et al. (2015) reported no significant impact of HELPP on body dissatisfaction at post-intervention, or follow-up. Despite WoE A penalties,

this study received a medium WoE D rating, with a high WoE B rating. This was due to its use of an RCT design, active comparison conditions, and power to detect small effect sizes.

McLean et al.'s (2019) study found no significant interaction of time and group ($\eta_p^2=.004$, $p>.05$), and post-hoc tests revealed a significant effect of time ($\eta_p^2=.014$, $p<.05$) but not group ($\eta_p^2=.001$, $p>.05$) on girls' body dissatisfaction. This suggests that participants in the HBM-Media and HBM-Comparison conditions did not experience significantly greater reductions in body dissatisfaction compared to the active control, HBM-Eating. McLean et al.'s (2019) study received a high WoE D rating, with its use of an RCT design and an active control condition contributing positively to WoE B ratings. Additionally, it was the only study to meet all essential WoE A criteria. The strong methodological quality and relevance of this study should be considered when interpreting findings, however, the study received WoE C penalties due to its dismantling of HBM into two separate intervention conditions.

Table 8

Summary of Effect Sizes for Reviewed Studies

Author	Measures	Total Sample Size (Female N if different)	Effect size pre-post η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	Effect size pre-follow up η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	WoE D
Richardson & Paxton (2010)	Extended Body Satisfaction Visual Analogue Scale ^a	194	.113***	Medium	+	.113***	Medium	+	2.1 (Medium)
	9-item Body Dissatisfaction subscale of the Eating Disorder Inventory ^b		.090***	Medium	+	.068**	Medium	+	
Bird et al. (2013)	Extended Body Satisfaction Visual	88 (42)	.30***	Large	+	.16*	Large	+	2.2 (Medium)

Author	Measures	Total Sample Size (Female N if different)	Effect size pre-post η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	Effect size pre-follow up η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	WoE D
	Analogue Scale ^a								
Stewart et al. (2020) (Comparison of intervention & Control)	Extended Body Satisfaction Analogue Scale ^a	369 (332)	.01*	Small	+	.02 (boys and girls only not available for follow-up)	Small	+ but no longer significant at follow-up	2.2 (Medium)
Dunstan et al. (2017)	9-item Eating Disorder Inventory – Body Dissatisfaction Subscale ^b	200	NA	NA	NA	.05*	Small	+ significant intervention effect observed at post intervention but not maintained at follow-up	2.5 (High)
Wilksch et al. (2015)	9-item Eating Disorder	1316 (840)	NA	NA	NA	NA	NA	-No significant interaction	2.2 (Medium)

Author	Measures	Total Sample Size (Female N if different)	Effect size pre-post η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	Effect size pre-follow up η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	WoE D
	Inventory – Body Dissatisfaction Subscale ^b							effects at post or follow-up	
McClellan et al. (2019)	9-item Eating Disorder Inventory – Body Dissatisfaction Subscale ^b	260	NA	NA	NA	.004	Small	-No significant interaction effects. Post-hoc revealed significant effect of time but not group.	2.6 (High)
Sub-section of Stewart et al. (2020) (<i>Comparison of clinician vs</i>	Extended Body Satisfaction Visual Analogue Scale ^a	346 (NA)	NA	NA	NA	.012*	Small	(Analysis not conducted separately for males	2.3 (Medium)

Author	Measures	Total Sample Size (Female N if different)	Effect size pre-post η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	Effect size pre-follow up η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	WoE D
<i>teacher delivery)</i>								and females) Post hoc tests revealed significant medium sized effect of HBM on body satisfaction in clinician-led group which was maintained at follow-up. Same effect not observed in teacher-led group.	

Author	Measures	Total Sample Size (Female N if different)	Effect size pre-post η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	Effect size pre-follow up η_p^2 (group x time interaction)	Descriptor	Finding (+positive impact of intervention, -negative or no impact)	WoE D
Sub-section of Dunstan et al. (2017) (<i>Comparison of co-educational versus single sex school</i>)	9-item Eating Disorder Inventory – Body Dissatisfaction Subscale ^b	200				.01	Small	-No significant group by time interactions indicating no significant differences between co-educational and single sex delivery	1.8 (Medium)

η_p^2 = partial eta squared.

^a Higher scores more desirable.

^b Lower scores more desirable.

NA: Information not available or could not be calculated from information provided in study.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Conclusions and Recommendations

This review evaluated the effectiveness of HBM in improving body satisfaction in pre-adolescent and adolescent females. In conclusion, evidence for the effectiveness of HBM is mixed, and various factors should be considered by EPs before recommending HBM.

Four studies found a significant, positive effect of HBM (Bird et al., 2013; Dunstan et al., 2017; Richardson & Paxton, 2010; Stewart et al., 2020), however, effect sizes varied. Importantly, three of the four studies which found a significant effect (Bird et al., 2013; Richardson & Paxton, 2010; Stewart et al., 2020), and specifically, the two studies which found the largest effects of HBM (Bird et al., 2013; Richardson & Paxton, 2010), were quasi-experimental in design and did not randomly allocate participants. One study allocated on a first-come, first-served basis (Richardson & Paxton, 2010), whilst in two studies, teachers allocated classes to conditions (Bird et al., 2013; Stewart et al., 2020). These methods increase the likelihood that systematic differences on unmeasured constructs may have caused improvements, rather than HBM. The positive findings of these three studies should therefore be interpreted with caution.

Additionally, two of the four studies reported that effects were no longer significant at follow-up (Dunstan et al., 2017; Stewart et al., 2020). It is unclear why the positive effects of HBM were not maintained. However, Dunstan et al. (2017) suggest that their longer follow-up period of six months, as opposed to the three-month follow-up used in other studies, may have enabled more opportunities for challenges to body satisfaction. They recommend booster sessions, to address needs at regular intervals, may be useful in maintaining the positive effects of HBM. This is a long-term factor EPs should consider if recommending HBM.

It is important to consider adaptations to HBM's original format when thinking about the variability in effects. Five studies made adaptations and only one study included measures of intervention fidelity (McLean et al., 2019). Bird et al. (2013), who found a large effect of

HBM, made minimal adaptations without increasing the number of sessions. McLean et al. (2019) also delivered three sessions, but dismantled the intervention, which may have negatively impacted the effectiveness of HBM. The remaining studies included additional sessions and content when implementing HBM. Whilst it is often appropriate to adapt interventions, particularly in schools, intervention fidelity and adaptation must be balanced, as sometimes, adapting interventions can reduce effectiveness (Toomey et al., 2020). For example, Wilksch et al. (2015) extended HBM to eight sessions and added content about eating concerns. However, the authors suggest that incorporating this may have drawn negative attention to eating behaviours. Therefore, thorough consideration of adaptations made to HBM will need to ensure that they correspond with the psychological theory, and aims of the original intervention. It may be tentatively suggested that based on findings within this review, three-session implementations of HBM might be more effective than extended versions. However, this must be considered in light of the aforementioned design limitations of studies.

Dunstan et al. (2017) found that delivering HBM in a co-educational setting was no less effective than delivery in a single-sex setting. This is a positive finding for the applicability of HBM to a UK educational context, where the majority of schools are co-educational (DfE, 2021). However, evidence that HBM delivered by teachers was less effective when compared to clinician delivery, raises concerns about the feasibility of school implementation (Stewart et al., 2020). Despite not utilising formal fidelity measures, Stewart et al. (2020) collected feedback from teachers, who found the intervention content repetitive, and based on time constraints, altered content. This finding reinforces the need for studies to measure intervention fidelity, and may provide support for the 3-session HBM, which could be more feasibly implemented, with less repetition. Stewart et al. (2020) found that teachers were motivated to deliver HBM, but wanted to incorporate their own adaptations, which may have led to a dilution of effectiveness. Therefore, if an EP was to recommend teacher delivery of

HBM, they should provide formal training and supervision throughout implementation, involving discussions of the importance of balancing fidelity and adaptations.

Limitations

The decision to focus on the effectiveness of HBM for girls in this review was based on evidence that this population is a high-risk group (Dunstan et al., 2017). However, males are increasingly reporting body dissatisfaction, and three studies in the review included males (Bird et al., 2013; Stewart et al., 2020; Wilksch et al., 2015). Taking a broader approach, including males, females, and those who may be gender questioning, may provide a richer and more inclusive insight into the effectiveness of HBM. It could also provide information about the feasibility of the intervention in co-educational settings, where single-gender groupings are difficult to implement. However, because only three studies included boys, it may have been difficult to make robust conclusions about the effectiveness for males.

Future Research

To inform EPs' recommendations about the implementation of HBM in schools, future research could compare the effectiveness of the original HBM with adapted versions. Additionally, future research should endeavour to conduct formal measures of fidelity. This may be particularly relevant when evaluating teacher delivery. For example, research could evaluate effectiveness and fidelity when teachers have received formalised training and supervision, compared to when they have not. This could contribute to greater understanding of the feasibility of implementing HBM in schools, and would help EPs to make informed recommendations about the best way of ensuring the programme's effectiveness.

References

- Barker, C., Pistrang, N., & Elliott, R. (2016). *Research methods in clinical psychology: An introduction for students and practitioners* (3rd ed.). Wiley-Blackwell.
- Bird, E. L., Halliwell, E., Diedrichs, P. C., & Harcourt, D. (2013). Happy being me in the UK: A controlled evaluation of a school-based body image intervention with pre-adolescent children. *Body Image*, *10*(3), 326–334.
<https://doi.org/10.1016/j.bodyim.2013.02.008>
- Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. *Perspectives on Psychological Science*, *8*(4), 445–454. <https://doi.org/10.1177/1745691613491271>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (revised ed.). Academic Press. New York.
- Department for Education. (2020, June 25). *Statistics: School and pupil numbers*.
<https://www.gov.uk/government/collections/statistics-school-and-pupil-numbers>
- Department of Health and Social Care and Department of Education. (2017, December). *Transforming Children and Young People’s Mental Health Provision: A Green Paper*. <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>
- Dunstan, C. J., Paxton, S. J., & McLean, S. A. (2017). An evaluation of a body image intervention in adolescent girls delivered in single-sex versus co-educational classroom settings. *Eating Behaviors*, *25*, 23–31.
<https://doi.org/10.1016/j.eatbeh.2016.03.016>

- Durkin, S. J., & Paxton, S. J. (2002). Predictors of vulnerability to reduced body image satisfaction and psychological wellbeing in response to exposure to idealized female media images in adolescent girls. *Journal of Psychosomatic Research, 53*(5), 995–1005. [https://doi.org/10.1016/S0022-3999\(02\)00489-0](https://doi.org/10.1016/S0022-3999(02)00489-0)
- Durkin, S. J., Paxton, S. J., & Sorbello, M. (2007). An Integrative Model of the Impact of Exposure to Idealized Female Images on Adolescent Girls? Body Satisfaction. *Journal of Applied Social Psychology, 37*(5), 1092–1117. <https://doi.org/10.1111/j.1559-1816.2007.00201.x>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Florin, T. A., Shults, J., & Stettler, N. (2011). Perception of overweight is associated with poor academic performance in US adolescents. *Journal of School Health, 81*(11), 663–670. <https://doi.org/10.1111/j.1746-1561.2011.00642.x>
- Garner, D. M., Olmstead, M. P., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders, 2*(2), 15–34. [https://doi.org/10.1002/1098-108X\(198321\)2:2<15::AID-EAT2260020203>3.0.CO;2-6](https://doi.org/10.1002/1098-108X(198321)2:2<15::AID-EAT2260020203>3.0.CO;2-6)
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164. <https://doi.org/10.1177/001440290507100202>

- 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>
- Government Equalities Office. (2013). *Body confidence campaign Progress report 2013*.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/203405/130501_body_confidence_progress_reportv03.pdf
- Halliwell, E., Diedrichs, P., & Orbach, S. (2014, December 3). *Costing the Invisible: A review of the evidence examining the links between body image, aspirations, education and workplace confidence*. <https://uwe-repository.worktribe.com/output/806655/costing-the-invisible-a-review-of-the-evidence-examining-the-links-between-body-image-aspirations-education-and-workplace-confidence>
- Hicks, T. B., Shahidullah, J. D., Carlson, J. S., & Palejwala, M. H. (2014). Nationally certified school psychologists' use and reported barriers to using evidence-based interventions in schools: The influence of graduate program training and education. *School Psychology Quarterly*, 29(4), 469–487.
<https://doi.org/10.1037/spq0000059>
- Jones, D. C. (2001). Social comparison and body image: Attractiveness comparisons to models and peers among adolescent girls and boys. *Sex Roles*, 45(9–10), 645–664. <https://doi.org/10.1023/A:1014815725852>
- Jones, D.C., Vigfusdottir, H., & Lee, Y. (2004). *Conversations, Peer Criticism, Appearance Magazines, and the Internalization of Appearance Ideals Body Image and the Appearance Culture Among Adolescent Girls and Boys: An*

examination of friend conversations, peer criticism, appearance magazines, and the internalization of appearance ideals. *Journal of adolescent research*, 19(3), 323-339 <https://doi.org/10.1177/0743558403258847>

Kratochwill, T. R. (2007). Preparing Psychologists for Evidence-Based School Practice: Lessons Learned and Challenges Ahead. *American Psychologist*, 62(8), 829–843. <https://doi.org/10.1037/0003-066X.62.8.829>

Lovegrove, E., & Rumsey, N. (2005). Ignoring it doesn't make it stop: Adolescents, appearance, and bullying. *Cleft Palate-Craniofacial Journal*, 42(1), 33–44. <https://doi.org/10.1597/03-097.5.1>

McLean, S. A., Wertheim, E. H., Marques, M. D., & Paxton, S. J. (2019). Dismantling prevention: Comparison of outcomes following media literacy and appearance comparison modules in a randomised controlled trial. *Journal of Health Psychology*, 24(6), 761–776. <https://doi.org/10.1177/1359105316678668>

Mikkilä, V., Lahti-Koski, M., Pietinen, P., Virtanen, S. M., & Rimpelä, M. (2003). Associates of obesity and weight dissatisfaction among Finnish adolescents. *Public Health Nutrition*, 6(1), 49–56. <https://doi.org/10.1079/phn2002352>

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, 57(7), 527-529. <https://doi.org/10.1136/jech.57.7.527>

Richardson, S. M., & Paxton, S. J. (2010). An evaluation of a body image intervention based on risk factors for body dissatisfaction: A controlled study with adolescent girls. *International Journal of Eating Disorders*, 43(2), 112–122. <https://doi.org/10.1002/eat.20682>

- Rodgers, R. F., Paxton, S. J., & McLean, S. A. (2014). A Biopsychosocial Model of Body Image Concerns and Disordered Eating in Early Adolescent Girls. *Journal of Youth and Adolescence*, 43(5), 814–823. <https://doi.org/10.1007/s10964-013-0013-7>
- Sands, E. R., & Wardle, J. (2003). Internalization of ideal body shapes in 9-12-year-old girls. *International Journal of Eating Disorders*, 33(2), 193–204. <https://doi.org/10.1002/eat.10121>
- Stewart, C., Goddard, E., Cakir, Z., Hall, R., & Allen, G. (2020). Can more people be “Happy Being Me”? Testing the delivery of a universal body satisfaction program by clinicians and school staff. *Eating Disorders*, 1–28. <https://doi.org/10.1080/10640266.2020.1771165>
- Toomey, E., Hardeman, W., Hankonen, N., Byrne, M., McSharry, J., Matvienko-Sikar, K., & Lorencatto, F. (2020). Focusing on fidelity: Narrative review and recommendations for improving intervention fidelity within trials of health behaviour change interventions. *Health Psychology and Behavioral Medicine*, 8(1), 132–151. <https://doi.org/10.1080/21642850.2020.1738935>
- Wilksch, S. M., Paxton, S. J., Byrne, S. M., Austin, S. B., McLean, S. A., Thompson, K. M., Dorairaj, K., & Wade, T. D. (2015). Prevention Across the Spectrum: A randomized controlled trial of three programs to reduce risk factors for both eating disorders and obesity. *Psychological Medicine*, 45(9), 1811–1823. <https://doi.org/10.1017/S003329171400289X>

Appendices

Appendix A: Articles Excluded at Full Text Screening

Articles excluded at full text screening	Excluded on criteria:
<p>Roberts-Parker, E. (2019). <i>Evaluating a Mentalization Based Program for Addressing Eating Disorder Risk Among Young Adolescents</i> (Doctoral dissertation, Alliant International University).</p>	<p>8- The study's control group also receives HBM with additional mentalisation component. 9- The study does not include a measure of body dissatisfaction/satisfaction.</p>
<p>McCabe, M. P., Connaughton, C., Tatangelo, G., Mellor, D., & Busija, L. (2017). Healthy me: A gender-specific program to address body image concerns and risk factors among preadolescents. <i>Body image</i>, 20, 20-30. https://doi.org/10.1016/j.bodyim.2016.10.007</p>	<p>1- Study did not evaluate effectiveness of HBM intervention 2- Study included participants under the age of 10</p>
<p>Wilksch, S. M., Paxton, S. J., Byrne, S. M., Austin, S. B., O'Shea, A., & Wade, T. D. (2017). Outcomes of three universal eating disorder risk reduction programs by participants with higher and lower baseline shape and weight concern. <i>International Journal of Eating Disorders</i>, 50(1), 66-75. https://doi.org/10.1002/eat.22642</p>	<p>6- The study does not contain primary empirical data</p>

Appendix B: Mapping the Field

Authors and Title	Sample Size	Sample characteristics	Study Design	Country	Intervention	Control	Body Dis/satisfaction Outcome Measure	Primary Findings
<p>Authors: Richardson and Paxton (2010)</p> <p>Title: An evaluation of a body image intervention based on risk factors for body dissatisfaction: A controlled study with adolescent girls</p>	<p>Total= 194 Treatment= 104 Control= 90</p>	<p>Female Mean age= 12 years 4 months (SD= 4.13 months)</p>	<p>Non-randomised group design 3-month follow-up</p>	<p>Australia</p>	<p>3, 50-minute sessions of Happy Being Me (HBM) during usual health lessons.</p>	<p>No-intervention control but wait list.</p>	<p>Extended version of Body Satisfaction Visual Analogue Scale 9-item Body dissatisfaction subscale of the Eating Disorder Inventory</p>	<p>Moderate-large effect of intervention on body dissatisfaction at post-intervention and follow-up.</p>
<p>Authors: Bird, Halliwell, Diedrichs and Harcourt (2013)</p> <p>Title: Happy Being Me in the UK: A controlled evaluation of a school-based body image intervention with pre-adolescent children</p>	<p>Total= 88 Treatment=43 Control= 45</p>	<p>46 Males, 42 Females Age= 10-11 years</p>	<p>Quasi-experimental design 3-month follow-up</p>	<p>England</p>	<p>3, 1-hour sessions with one week between each session. Adapted HBM for delivery with males and females and for age range.</p>	<p>No-intervention control.</p>	<p>Adapted version of Body Satisfaction Visual Analogue Scale</p>	<p>Time x condition interactions were significant for girls' body satisfaction with improvement in intervention group at post-intervention and follow up.</p>

Authors and Title	Sample Size	Sample characteristics	Study Design	Country	Intervention	Control	Body Dis/satisfaction Outcome Measure	Primary Findings
<p>Authors: Stewart, Goddard, Cakir, Hall and Allen (2020)</p> <p>Title: Can more people be “Happy Being Me”? Testing the delivery of a universal body satisfaction program by clinicians and school staff</p> <p><i>(Study 1 compares HBM and comparison group, Study 2 compares clinician versus school staff delivery)</i></p>	<p>Total= 369 Treatment= 172 Control= 197</p>	<p>Year 7 boys (N=37) and girls (N=332) (aged 11–12 years) recruited from 8 schools</p> <p>Mean age= 11.7 SD= 0.33</p>	<p>Quasi-experimental case-controlled comparison trial</p>	<p>UK</p>	<p>6, weekly, 50-minute HBM sessions delivered in PSHE lessons.</p>	<p>No-intervention control.</p>	<p>Extended version of the Body Satisfaction Visual Analogue Scale (BSVA; Durkin & Paxton, 2002)</p>	<p>Improvement in body satisfaction in HBM group for girls.</p> <p>Follow-up data not stratified by gender but no longer significant at follow-up.</p> <p>Comparison of teacher versus clinician delivery found improvements in clinician delivery but not teacher.</p>

Authors and Title	Sample Size	Sample characteristics	Study Design	Country	Intervention	Control	Body Dis/satisfaction Outcome Measure	Primary Findings
<p>Authors: Dunstan, Paxton and McLean (2017)</p> <p>Title: An evaluation of a body image intervention in adolescent girls delivered in single-sex versus co-educational classroom settings</p> <p><i>(Additional part of study compares co-educational and single-sex outcomes)</i></p>	<p>Total = 200 Co-educational delivery= 73 Single-sex delivery= 74 No intervention control= 53</p>	<p>Grade 7 girls aged between 11-14.</p>	<p>Block RCT</p> <p>Comparison of co-educational/single-sex delivery & no intervention control</p> <p>6- month follow-up</p>	<p>Australia</p>	<p>6, weekly sessions of HBM</p>	<p>No-intervention control.</p>	<p>9-item Eating Disorder Inventory – Body Dissatisfaction Subscale</p>	<p>Significantly lower body dissatisfaction at post-intervention in HBM group but not maintained at follow-up.</p> <p>No significant difference between co-educational delivery and single-sex delivery.</p>

Authors and Title	Sample Size	Sample characteristics	Study Design	Country	Intervention	Control	Body Dis/satisfaction Outcome Measure	Primary Findings
<p>Authors: Wilksch, Paxton, Byrne, Austin, McLean, Thompson, Dorairaj, & Wade (2015)</p> <p>Title: Prevention Across the Spectrum: A randomized controlled trial of three programs to reduce risk factors for both eating disorders and obesity</p>	Total= 1316 Girls N= 840 Boys N= 476	Grade 7 boys and girls Mean age= 13.21 (SD= 0.68)	Randomised controlled trial (class) Comparison of 3 different school-based body image interventions with a no intervention control. 6 and 12-month follow-up	Australia	HBM extended to 8 sessions and including components on eating concerns. Re-named HELPP (Helping, Encouraging, Listening and Protecting Peers Initiative)	Usual school class/no-intervention control.	Eating disorder inventory body dissatisfaction 9-items.	No impact of intervention on body dissatisfaction.
<p>Authors: McLean, Wertheim, Marques, & Paxton (2016)</p> <p>Title: Dismantling prevention: Comparison of outcomes following media literacy and</p>	N= 260	Adolescent girls Aged 11-14	A three-arm cluster randomised controlled trial design with equal allocation.	Australia	HBM-Media and HBM-Comparison were adapted from previous versions of Happy Being Me (3 session versions)	Active control= HBM Healthy eating behaviour	Eating disorder inventory 9-item body dissatisfaction	Significant main effect of time but no interaction (group x time) on body dissatisfaction.

Authors and Title	Sample Size	Sample characteristics	Study Design	Country	Intervention	Control	Body Dis/satisfaction Outcome Measure	Primary Findings
appearance comparison modules in a randomised controlled trial								

Appendix C: Criteria and Rationale for WoE Ratings

WoE A: Methodological Quality

WoE A is described as a non-review specific judgement of the methodological quality of studies (Gough, 2007). As all reviewed studies employed a group experimental design, Gersten et al.'s (2005) coding protocol was considered appropriate in assessing WoE A. Adaptations were made to the protocol to ensure relevance to the review, and the adapted protocol and completed protocols for each study can be found in Appendix D and E respectively. The protocol includes both essential and desirable quality indicators and Table C1 indicates the cut-off points used to assess WoE A. The criteria for 'high' and 'medium' rated studies was based on those suggested by Gersten et al. (2005) but adapted due to the removal of one essential criteria from the protocol. The criteria for 'low' and 'very low' were not outlined by Gersten et al. (2005) but decided upon by the author for the purpose of the review. WoE A ratings for the six reviewed studies are provided in Table C2.

Table C1

Cut-off criteria for WoE A Rating

WoA Rating		Criteria
3	High	Meets at least 8 essential and 5 desirable criteria.
2	Medium	Meets at least 8 essential and between 1-4 desirable criteria.
1	Low	Meets at least 5 but lower than 8 essential criteria <i>or</i> does not meet any desirable criteria.
0	Very Low	Meets fewer than 5 essential criteria.

Table C2

WoE A ratings for reviewed studies

Summary of WoE A Ratings						
	Richardson and Paxton (2010)	Bird et al. (2013)	Stewart et al. (2020)	Dunstan et al. (2017)	Wilksch et al. (2015)	McClellan et al. (2016)
Essential criteria met /9	8	8	8	8	8	9
Desirable criteria met /11	7	5	4	5	4	7
WoE A Rating	3 (High)	3 (High)	2 (Medium)	3 (High)	2 (Medium)	3 (High)

WoE B: Methodological Relevance

WoE B is considered a review-specific judgement related to the appropriateness of the study methodology in answering the review question (Gough, 2007). The criteria for WoE B were compiled by the author and are presented in Table C3. The four criteria considered were study design, participant allocation, control condition, and study power, and the rationale for including these criteria is presented within Table C3. The WoE B ratings for the six reviewed studies can be found in Table C4. The overall WoE B rating for each study was calculated by averaging the scores across the four criteria. For the fourth WoE B criteria ‘Power’, the software G*Power (Faul et al., 2007) was used to calculate whether the study was appropriately powered to detect small, medium, and large effects. The input parameters utilised varied depending on the statistical analysis conducted, but each study was assessed

according to whether it was appropriately powered to detect a large, medium and small effect size. The respective input values suggested by G*Power for small, medium and large effects are outlined in Table C5.

Table C3

Criteria for WoE B with Rationale

Criteria	Low: 1	Medium: 2	High: 3	Rationale
Study Design	Cohort study design.	Quasi-experimental design.	Randomised controlled trial design.	RCTs are considered more methodologically appropriate than quasi-experimental and cohort studies for investigating effectiveness (Petticrew & Roberts, 2003).
Participant allocation	No random allocation of participants at individual or group level.	Random allocation of groups/block randomisation: e.g. classes or schools to intervention or control condition.	Random allocation of individuals to intervention or control condition.	Randomisation of participants limits the potential for selection bias and reduces the effects of pre-existing group differences (Barker et al., 2016).
Control condition	Participants receive no alternative intervention i.e.	Participants in control condition receive an alternative body	Participants in control condition receive HBM intervention	Active control groups increase confidence in conclusions that

Criteria	Low: 1	Medium: 2	High: 3	Rationale
	treatment as usual.	image intervention.	without active components 'media literacy' and 'appearance comparison'.	outcomes are not due to factors other than the intervention itself e.g. attention of facilitator (Boot et al., 2013).
Power	Sample size is not appropriate for the statistical analyses.	Sample size is appropriate to detect medium or large effect size.	Sample size is appropriate to detect small effect size.	Studies need to be appropriately powered in order to detect effects of the intervention (Barker et al., 2016).

Table C4

WoE B Ratings for Reviewed Studies

Summary of WoE B Ratings						
Criteria	Richardson & Paxton (2010)	Bird et al. (2013)	Stewart et al. (2020)	Dunstan et al. (2017)	Wilksch et al. (2015)	McClellan et al. (2016)
Study design	2	2	2	3	3	3
Participant allocation	1	1	1	2	2	2
Control condition	1	1	1	1	2	3
Power	2	2	3	3	3	3

WoE B	1.5	1.5	1.75	2.25	2.5	2.75
Rating	(Medium)	(Medium)	(Medium)	(Medium)	(High)	(High)

Note. WoE B ratings are described as ‘High’ for scores ≥ 2.5 , ‘Medium’ for scores ≥ 1.5 and < 2.5 , and ‘Low’ for scores < 1.5 .

Table C5

*G*Power Effect Size Input*

Effect Size f	Label
0.10	Small
0.25	Medium
0.40	Large

WoE C: Topic Relevance

WoE C is considered a review-specific judgement regarding the relevance of the topic of the study in answering the specific review question (Gough, 2007). The WoE C coding protocol was designed by the author. The criteria along with the rationale for these criteria, are presented in Table C6. WoE C ratings for the six reviewed studies can be found in Table C7. The overall WoE C rating for each study was calculated by averaging scores across the five criteria.

Table C6

Criteria for WoE C with Rationale

Criteria	Low: 1	Medium: 2	High: 3	Rationale
Location	Study is conducted in a location not economically similar to the UK.	Study is conducted in a location economically similar to the UK.	Study is conducted in the UK.	It is important that the study has taken place in a location with a comparable education system or it may not be possible to generalise

Criteria	Low: 1	Medium: 2	High: 3	Rationale
				the results to an English school setting.
Intervention delivery	Intervention is only delivered by external professional e.g. researcher.	Intervention is partially delivered by school staff and partially by an external professional.	Intervention is delivered by school staff or there is a comparison of delivery by school staff and researcher.	As the intervention is designed to be delivered in schools, a study which utilises school staff is more reflective of how the intervention will be implemented in practice.
Intervention implementation	Intervention delivered does not include the key components of the original HBM intervention.	Intervention delivered is an adapted version of original (either by increasing/ amending content or by dismantling intervention into component parts).	Intervention delivered is the original HBM programme or a replication of the original HBM intervention.	Studies which use the HBM intervention in its originally designed format answer the review question most appropriately. However, it is useful to include adaptations to understand whether these adaptations improve or reduce effectiveness.
Setting generalisability: number of schools	Participants are sampled from a single school.	Participants are sampled from two schools.	Participants are sampled from more than two schools.	Findings are more generalisable and more representative if they are based on multiple schools rather than one.
Setting generalisability:	Participants are only sampled	Participants are only sampled	Participants are sampled from a	Studies which only investigate the

Criteria	Low: 1	Medium: 2	High: 3	Rationale
co-educational or single sex setting	from single sex schools or information is not available.	from a co-educational school.	combination of single sex and co-educational schools.	effectiveness of HBM for girls who attend single sex schools cannot provide results which generalise to girls attending co-educational settings or evaluate whether the intervention delivered in one setting is more effective than the other.

Table C7

WoE C Ratings for Reviewed Studies

Summary of WoE C Ratings						
Criteria	Richardson & Paxton (2010)	Bird et al. (2013)	Stewart et al. (2020)	Dunstan et al. (2017)	Wilksch et al. (2015)	McClellan et al. (2016)
Location	2	3	3	2	2	2
Intervention delivery	1	1	3	1	1	1
Intervention implementation	3	2	2	2	2	2
Setting generalisability: number of schools	2	2	3	3	3	3

Summary of WoE C Ratings						
Criteria	Richardson & Paxton (2010)	Bird et al. (2013)	Stewart et al. (2020)	Dunstan et al. (2017)	Wilksch et al. (2015)	McClellan et al. (2016)
Setting generalisability: single sex or co-educational	1	2	3	3	3	1
WoE C Rating	1.8 (Medium)	2 (Medium)	2.8 (High)	2.2 (Medium)	2.2 (Medium)	1.8 (Medium)

Note. WoE C ratings are described as ‘High’ for scores ≥ 2.5 , ‘Medium’ for scores ≥ 1.5 and < 2.5 , and ‘Low’ for scores < 1.5 .

WoE D: Overall WoE Judgement

WoE D is an overall judgement of the quality and relevance of the studies reviewed calculated by taking an average across the three WoE (A, B, C). A summary of the WoE D ratings can be found in Table C8.

Table C8

WoE D Ratings for Reviewed Studies

Study	WoE A	WoE B	WoE C	WoE D (1 Decimal Place)
Richardson & Paxton (2010)	3 (High)	1.5 (Medium)	1.8 (Medium)	2.1 (Medium)
Bird et al. (2013)	3 (High)	1.5 (Medium)	2 (Medium)	2.2 (Medium)
Stewart et al. (2020)	2 (Medium)	1.75 (Medium)	2.8 (High)	2.2 (Medium)

Study	WoE A	WoE B	WoE C	WoE D (1 Decimal Place)
Dunstan et al. (2017)	3 (High)	2.25 (Medium)	2.2 (Medium)	2.5 (High)
Wilksch et al. (2015)	2 (Medium)	2.5 (High)	2.2 (Medium)	2.2 (Medium)
McClean et al. (2016)	3 (High)	2.75 (High)	1.8 (Medium)	2.5 (High)

Note. WoE D ratings are described as 'High' for scores ≥ 2.5 , 'Medium' for scores ≥ 1.5 and < 2.5 , and 'Low' for scores < 1.5 .

Appendix D: Adapted WoE A Coding Protocol (Gersten et al., 2005)

In order to ensure clarity and appropriateness of the coding protocol to the review, adaptations were made to Gersten et al.'s (2005) coding protocol. The striking through of text indicates where text has been removed, {} indicates text has been added, and the rationale for changes made is written in a coloured font beneath each item. The adapted protocol resulted in nine essential criteria and eleven desirable criteria, and cut-off points for coding were provided in Appendix C.

Essential Quality Indicators

Quality Indicators for Describing Participants

- ~~1. Was sufficient information provided to determine/confirm whether the participants demonstrated the disability(ies) or difficulties presented?~~

(This quality indicator was removed as it was not relevant to the review question. The current review question focuses on an intervention which can be implemented for both high risk groups and universally with pupils who do/do not have disabilities/difficulties.)

1. Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?

(For the purpose of this review, appropriate procedures will include both random allocation and statistical procedures to ensure equivalence of groups. Studies which use random allocation will then be appraised as having greater methodological relevance to the research question in WoE B.)

2. Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?

Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions

1. Was the intervention clearly described and specified? {If the intervention implemented was an adaptation of the original intervention, was sufficient information provided about the adaptations made?}

(An additional part was added to this quality indicator, as many studies included in the current review adapt the original intervention e.g. extending the number of intervention sessions)

2. Was the fidelity of implementation described and assessed?
3. Was the nature of services provided in comparison conditions described?

Quality Indicators for Outcome Measures

1. ~~Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance? {In addition to the primary outcome measure, were other relevant body image outcome measures utilised?}~~

This question was changed as the intervention which is the focus of the review question aims to target body image related concerns and does not aim to directly impact upon generalised performance. An additional desirable criterion has been added to assess whether studies use multiple measures for the primary outcome of interest in the current review: 'body dissatisfaction'.

2. Were outcomes for capturing the interventions effect measured at the appropriate times?

Quality Indicators for Data Analysis

1. Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?
2. Did the research report include not only inferential statistics but also effect size calculations?

Desirable Quality Indicators

1. {Did the study use more than one outcome measure related to the primary outcome of interest: body dissatisfaction?}

This question was added into the desirable indicators to account for the fact that some studies use more than one measure of the primary outcome measure of interest: 'body dissatisfaction'. E.g. some use a measure of body dissatisfaction alongside a measure of body satisfaction.

2. Was data available on attrition rates among intervention samples? ~~Was severe overall attrition documented? If so, is attrition comparable across samples? Is overall attrition less than 30%?~~

(Split this question in case part of the question required a 'yes' and others required a 'no'.)

3. Was severe overall attrition (over 30%) avoided? Is attrition comparable across samples?

(This is the second part of the original question, as a separate question for clarity).

4. Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures? ~~Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?~~

(Question seems to be addressing two different issues, so I decided to split this question into two).

5. Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?
6. Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?

7. Was evidence of the criterion-related validity and construct validity of the measures provided?
8. Did the research team assess not only surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?
9. Was any documentation of the nature of instruction or series provided in comparison conditions?
10. Did the research report include actual audio or videotape excerpts {or examples of content} that capture the nature of the intervention?

(Many papers include a session guide including an outline of the content of the sessions which could be deemed as appropriate evidence).

11. Were results presented in a clear, coherent fashion?

Appendix E: WoE A Coded Study

Paper reference: Richardson, S. M., & Paxton, S. J. (2010). An evaluation of a body image intervention based on risk factors for body dissatisfaction: A controlled study with adolescent girls. *International Journal of Eating Disorders*, 43(2), 112-122.

Essential Quality Indicators

Quality Indicators for Describing Participants

2. Were appropriate procedures used to increase the likelihood that relevant characteristics of participants in the sample were comparable across conditions?
 - Yes- Analyses were conducted to assess if there were any differences between groups (intervention vs. control) at T1. No significant differences were found.
 - No
 - Unknown/unable to code

3. Was sufficient information given characterizing the interventionists or teachers provided? Did it indicate whether they were comparable across conditions?
 - Yes- intervention facilitated by first author.
 - No
 - Unknown/unable to code

Quality Indicators for Implementation of the Intervention and Description of Comparison Conditions

4. Was the intervention clearly described? If the intervention implemented was an adaptation of the original intervention, was sufficient information provided about the adaptations made
 - Yes
 - No
 - Unknown/unable to code

5. Was the fidelity of implementation described and assessed?

- Yes
- No
- Unknown/unable to code

6. Was the nature of services provided in comparison conditions described?

- Yes
- No
- Unknown/unable to code

Quality Indicators for Outcome Measures

3. In addition to the primary outcome measure, were other relevant body image outcome measures utilised?

- Yes
- No
- Unknown/unable to code

4. Were outcomes for capturing the interventions effect measured at the appropriate times?

- Yes
- No
- Unknown/unable to code

Quality Indicators for Data Analysis

3. Were the data analysis techniques appropriately linked to key research questions and hypotheses? Were they appropriately linked to the unit of analysis in the study?

- Yes
- No
- Unknown/unable to code

4. Did the research report include not only inferential statistics but also effect size calculations?

- Yes
- No
- Unknown/unable to code

Desirable Quality Indicators

1. Did the study use more than one outcome measure related to the primary outcome of interest: body dissatisfaction?

- Yes
- No
- Unknown/unable to code

2. Was data available on attrition rates among intervention samples?

- Yes
- No
- Unknown/unable to code

3. Was severe overall attrition (over 30%) avoided? Is attrition comparable across samples?

- Yes
- No

Unknown/unable to code

4. Did the study provide not only internal consistency reliability but also test-retest reliability and interrater reliability (when appropriate) for outcome measures?

Yes

No

Unknown/unable to code

5. Were data collectors and/or scorers blind to study conditions and equally (un)familiar to examinees across study conditions?

Yes

No

Unknown/unable to code

6. Were outcomes for capturing the intervention's effect measured beyond an immediate post-test?

Yes

No

Unknown/unable to code

7. Was evidence of the criterion-related validity and construct validity of the measures provided?

Yes

No

Unknown/unable to code

8. Did the research team assess not only surface features of fidelity implementation (e.g., number of minutes allocated to the intervention or teacher/interventionist following procedures specified), but also examine quality of implementation?

- Yes
- No
- Unknown/unable to code

9. Was any documentation of the nature of instruction or series provided in comparison conditions?

- Yes
- No- just says 'usual classes' with no additional information.
- Unknown/unable to code

10. Did the research report include actual audio, or videotape excerpts, or examples of content, that capture the nature of the intervention?

- Yes
- No
- Unknown/unable to code

11. Were results presented in a clear, coherent fashion?

- Yes
- No
- Unknown/unable to code

Essential criteria met: 8/9

Desirable criteria met: 7/11

Appendix F: Descriptive Statistics

Author	N (Females Only)	Measures	Intervention			Control		
			Pre Mean (SD)	Post Mean (SD)	Follow- up Mean (SD)	Pre Mean (SD)	Post Mean (SD)	Follow- up Mean (SD)
Richardson & Paxton (2010)	194	Extended Body Satisfaction Visual Analogue Scale ^a	322.18 (119.21)	365.51 (114.18)	386.07 (97.17)	281.43 (125.20)	280.35 (127.47)	290.83 (134.66)
		9-item Body Dissatisfaction subscale of the Eating Disorder Inventory ^b	24.19 (10.73)	21.09 (9.94)	19.64 (8.54)	26.52 (10.88)	26.67 (10.08)	26.26 (10.67)
Bird et al. (2013)	42	Extended Body Satisfaction Visual Analogue Scale ^a	11.60 (3.03)	13.63 (2.93)	15.09 (3.21)	12.77 (3.85)	11.1 (3.87)	13.64 (1.91)
Stewart et al. (2020) (Comparison of intervention & Control)	332	Extended Body Satisfaction Visual Analogue Scale ^a	345.4 (118.9)	372.3 (111.6)	358.9 (120.6)	355.7 (105.8)	355.6 (105.7)	348.8 (104.1)
Dunstan et al. (2017)	200	9-item Eating Disorder Inventory – Body Dissatisfaction Subscale ^b	28.35 (9.94)	26.09 (10.31)	28.19 (10.54)	24.40 (9.09)	26.37 (10.26)	27.17 (10.76)

Author	N (Females Only)	Measures	Intervention			Control		
			Pre Mean (SD)	Post Mean (SD)	Follow- up Mean (SD)	Pre Mean (SD)	Post Mean (SD)	Follow- up Mean (SD)
Wilksch et al. (2015)	840	9-item Eating Disorder Inventory – Body Dissatisfaction Subscale ^b	NA	1.89 (.10)*	2.04 (.10)*	NA	1.88 (.07)*	2.06 (.08)
<i>Sub-section of Stewart et al. (2020)</i>	N (did not analyse for boys/girls separately)	Measure	Pre Teacher Mean (SD)	Post Teacher Mean (SD)	Follow- up Teacher Mean (SD)	Pre Clinician Mean (SD)	Post Clinician Mean (SD)	Follow- up Clinician Mean (SD)
<i>(Comparison of clinician vs teacher delivery)</i>	346	Extended Body Satisfaction Visual Analogue Scale ^a	346.0 (113.9)	346.7 (124.0)	346.3 (116.3)	345.4 (118.9)	372.3 (111.6)	358.9 (120.6)
<i>Sub-section of Dunstan et al. (2017)</i>	N	Measure	Pre co- ed Mean (SD)	Post co- ed Mean (SD)	Follow- up co- ed Mean (SD)	Pre single Mean (SD)	Post single Mean (SD)	Follow- up single Mean (SD)
<i>(Comparison of co-educational versus single sex school)</i>	200	9-item Eating Disorder Inventory – Body Dissatisfaction Subscale ^b	28.28 (10.06)	26.69 (10.87)	29.13 (11.01)	28.41 (9.90)	25.50 (9.77)	27.26 (10.05)

Author	N (Females only)	Measure	HBM- media			HBM-comparison			HBM- eating (active control)		
			Pre Mean (SD)	Post Mean (SD)	Follow-up Mean (SD)	Pre Mean (SD)	Post Mean (SD)	Follow-up Mean (SD)	Pre Mean (SD)	Post Mean (SD)	Follow-up Mean (SD)
McClellan et al. (2016)	260	9-item Eating Disorder Inventory – Body Dissatisfaction Subscale ^b	29.64 (10.59)	30.00 (11.69)	29.43 (11.30)	28.64 (10.44)	28.99 (9.76)	28.27 (8.58)	26.57 (10.30)	26.48 (10.93)	27.44 (9.70)

(*)= SEM

^a= Higher scores more desirable

^b= Lower scores more desirable

NA= Not available