

Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

How effective is the Pyramid Club Intervention in improving the social-emotional wellbeing of primary and elementary school-aged pupils?

Summary

The Pyramid Club Intervention is a targeted small-group intervention for children who demonstrate internalising behaviours or are experiencing difficulties with peer relationships. It aims to give children the opportunity to build confidence, develop social skills and learn strategies to use to manage their thoughts and emotions (ContinYou, 2017; UWL, 2017a).

The aim of this systematic literature review is to evaluate the effectiveness of the Pyramid Club Intervention in improving the social-emotional wellbeing of primary and elementary school-aged pupils. A systematic literature review was undertaken using four online databases (PsychINFO, Web of Science, ERIC (EBSCO) and Google Scholar). Five studies were selected for review and were subsequently evaluated using Gough's (2007) Weight of Evidence Framework and the APA Task Force Coding Protocol by Kratchowill (2003). All studies used group-based designs and included outcome measures for pre-intervention and post-intervention. Three studies also included a follow-up outcome measure at 12 weeks.

The studies showed small to large effects at post-intervention and follow-up for all outcome measures reviewed. The largest effects were reported for the reduction in Emotional Difficulties, Peer Problems and Total Difficulties scores at follow-up and for improvements in Emotional Intelligence scores at both post-intervention and follow-up.

The implications of the review findings, as well as its limitations and future research suggestions are also discussed.

Introduction

What is Pyramid Club?

The Pyramid Club intervention is a targeted, small group therapeutic intervention aimed at children aged 7-14 years who are demonstrating internalising behaviours (e.g. shy, withdrawn or isolated) or early indicators of internalising mental health difficulties (e.g. depression, social withdrawal and isolation or anxiety; ContinYou, 2017; UWL, 2017a). It also targets children who are experiencing difficulties with peer relationships. It was developed in the UK during the 1970s (UWL, 2017a). It is usually implemented in Primary Schools either at the transition to Key Stage 2 (Years 3 and 4) or at the end of Primary School education (Years 5 and 6), which is the focus for this review. It can also be implemented in Secondary Schools (ContinYou, 2017). It adopts a preventative approach, which aims to target maladaptive internalising behaviours or emotions early on to prevent the development of more serious mental health conditions later on in life (ContinYou, 2017).

What it involves

The Pyramid Club is a manualised intervention which is delivered by trained Pyramid Club volunteers. They aim to create a supportive environment for the children and adapt activities to meet their needs (UWL, 2017a). It is designed to run as a weekly after-school club. Each session is 90 minutes long. The intervention is typically implemented for 10 weeks (UWL, 2017a). A range of activities are used to develop the children's social skills, confidence, emotional resilience and well-being. The main elements of the Pyramid Club Intervention model include naming and creating a set of rules for the group during the first week. In subsequent sessions the format

includes a circle time, art and physical activity e.g. team building games, group snack time and a closing circle time (UWL, 2017b; Ohl, Mitchell, Cassidy & Fox, 2008).

Key features of the Pyramid Club intervention model

The Pyramid Club intervention follows a three-stage selection and implementation model (Ohl, Mitchell, Cassidy & Fox, 2008). This is detailed in Table 1.

Table 1: *The Key Features of the Pyramid Club Intervention Model*

Stage	Description
1 Whole class screening	<p>Class teachers assess the children’s social and emotional wellbeing using the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; 1999).</p> <p>The scores are considered alongside the teacher’s knowledge of each child to guide selection of children for this intervention. The selection procedures outlined in the Pyramid manual (2007) also guide this process.</p>
2 Multi-agency meeting	<p>This meeting involves a discussion about the children who were identified as a potential cause for concern during stage one.</p> <p>A range of professionals will attend this meeting including:</p> <p>School staff e.g. Class Teachers, Head Teachers and the Special Educational Needs Coordinator (SENCo)</p> <p>A local Pyramid Co-ordinator</p> <p>Other relevant professionals linked to the children</p>

Stage	Description
3	Information provided by the Class Teachers, the results from the SDQ and other relevant information is used to select children to attend the Pyramid Club or receive another form of intervention.
Implementation	Children’s progress is monitored throughout.

How it helps

According to ContinYou (2017), the Pyramid Club allows the target children to become part of a small, supportive group where they can build positive relationships and experience fun learning experiences. Having these positive experiences may in turn help to encourage the targeted children to participate in learning activities in school and therefore have a positive impact on their attainment levels. ContinYou (2017) have identified many positive outcomes associated with attendance to this intervention including an increase in self-esteem, resilience and emotional regulation, more positive relationships with adults and peers and improvements in school achievement.

Psychological basis for Pyramid Club

Research findings demonstrating the potential success of small-group therapeutic interventions and the importance of peer relationships for children’s development at this stage contributed to the development of the Pyramid Club intervention (Lyons, 2011; ContinYou, 2017; UWL, 2017a).

Two main psychological models underpin the principles of the Pyramid Club (ContinYou, 2017; UWL, 2017a). The first is Positive Psychology, which focuses on

positive emotions and individual character traits and aims to support individuals to develop to their full potential and thrive. Interventions based on this model intend to increase individual happiness and enhance experiences of positive emotions (Seligman, Steen, Park & Peterson, 2005).

The second is Cognitive Psychology. The Pyramid Club supports children to learn how to manage their thoughts and feelings (ContinYou, 2017; UWL, 2017a). The Club utilises Kolb's experiential learning cycle (1984, as cited in Lyons, 2011). This gives children the opportunity to learn and develop new skills through their own direct experiences. This supports the application of skills in the long-term to other everyday situations outside of the intervention (Lyons, 2011).

Reference is also made to Maslow's Hierarchy of Needs in the Pyramid Club's Manual (2007; as cited in Lyons, 2011). The manual highlights how the experiences offered in Pyramid Clubs meet these needs in a variety of ways to promote the children's social and emotional development. For example, sharing snacks and drinks, having a clear routine to follow and opportunities to develop new friendships and become part of a group helps to meet the physiological, security, love and belonging and esteem needs of the children (Lyons, 2011).

Rationale for review

Children and young people's mental health and wellbeing is a major area of concern at present, with as many as one in ten children requiring a form of support or intervention for their difficulties (Department for Health, 2015). Several mental health difficulties begin to develop in childhood (Department for Health, 2015). Many children who do not receive early support are likely to experience persistent emotional or mental health difficulties into later life (Dunsmuir & Cobbald, 2017).

Teachers can find it more difficult to identify children with internalising disorders (e.g. anxiety or depression) compared to those who may have externalising disorders (e.g. aggression or anti-social behaviour; Dunsmuir & Cobbald, 2017). Children with internalising disorders may also find it more difficult to communicate to others how they are feeling (Dunsmuir & Cobbald, 2017). As a result, fewer children with internalising disorders are identified and therefore do not receive appropriate support.

At present, there are growing concerns about children and young people who display anxious and withdrawn behaviours as these behaviours are associated with a variety of poorer outcomes (Department for Education, 2016). For example, research has indicated that children who appear shy and withdrawn are at a higher risk of experiencing low self-esteem and confidence, anxiety, depression and difficulties with peer and teacher interactions and relationships (Rubin, Coplan & Bowker, 2009). Additionally, Fitzherbert (1997) noted that children with low self-esteem may appear quiet and withdrawn, and highlighted that children who experience early emotional difficulties may be at a higher risk of school exclusion. Therefore, it is important for schools to be able to effectively identify and support this group of children who may be at risk of poorer outcomes, as well as social and school exclusion.

Schools are well placed to identify children's social-emotional and wellbeing needs and intervene early. The recent SEND Code of Practice (DfE & DoH, 2015) emphasised the need to implement early intervention and preventative approaches to reduce the need for more specialist intervention at later stages. As highlighted by Frederickson and Cline (2015), schools benefit hugely from Educational Psychology involvement when working to promote the social, emotional and mental health

wellbeing of all pupils. For example, Educational Psychologists (EPs) are able to offer guidance to schools when selecting appropriate targeted interventions.

To be able to do this effectively, EPs need to have knowledge of the efficacy of a range of interventions. This also relates to the SEND Code of Practice (DfE & DoH, 2015), which has highlighted the need for evidence-based interventions to be implemented to support and meet children's needs effectively. This systematic review aims to contribute to EP practice in this way and provide an independent review of the literature into the effectiveness of the Pyramid Club intervention. It is hoped that this will help to guide decision-making when recommending suitable support for groups of children within school settings.

Review question

How effective is the Pyramid Club Intervention Model in improving the social-emotional wellbeing of primary and elementary school aged pupils?

Critical Review of the Evidence Base for Pyramid Club Intervention

Literature Search

A literature search was conducted between 8th January and 4th February 2017 using various online databases including PsychINFO, Web of Science, ERIC (EBSCO) and Google Scholar. To identify potential studies to include in this literature review, the reviewer selected the following search terms (see Table 2).

Table 2: *Online Database Search Terms*

Databases searched	Search terms
PsychINFO	“pyramid club”
Web of Science	AND
ERIC (EBSCO)	“school based intervention OR intervention”
Google Scholar	AND “primary school OR elementary school”

103 studies were found. Six papers were duplicates. The reviewer screened the studies by title using the inclusion and exclusion criteria (see Table 3). This excluded 91 studies from the review. After abstract screening, six studies were brought forward for full text screening. One study was subsequently excluded from the review (see Appendix A for rationale). This left a total of five studies to include in this review. An ancestral search of these papers revealed two possible studies for review, which were later excluded (see Appendix A for rationale). Figure 1 illustrates the process of the literature search completed.

Table 3: *Inclusion and Exclusion Criteria Used to Select Studies for the Review*

Criterion	Criteria for inclusion	Criteria for exclusion	Rationale
1. Publication Type	The study has been published in a peer-reviewed journal.	The study has not been published in a peer-reviewed journal.	Studies included within peer-reviewed journals are likely to have implemented a higher standard of research design and scientific rigour.
2. Language	The study has been published in the English language.	The study has not been published in the English language.	To ensure that the reviewer can obtain a solid understanding of the paper.
3. Research Design and methodology	<p>The study employed an experimental design, with primary empirical data collected at pre and post intervention.</p> <p>The outcome measures used must focus on social and emotional wellbeing.</p>	<p>The study did not employ an experimental design. Primary empirical data was not collected at pre and post intervention.</p> <p>The outcome measures used do not measure social and emotional wellbeing.</p>	<p>(a) To ensure that original data is reviewed.</p> <p>(b) Quantitative data is required to critically evaluate the effectiveness of the intervention and allow the reviewer to calculate effect sizes to compare results across studies.</p> <p>(c) To review the effectiveness of the intervention on a child's social, emotional and mental health wellbeing.</p>
4. Type of Intervention	The study must have delivered the Pyramid Club intervention.	The study did not deliver the Pyramid Club intervention.	To enable the reviewer to critically evaluate research on the Pyramid Club intervention.
5. Data analysis	<p>The study must report the following:</p> <ol style="list-style-type: none"> 1) Means 2) Standard deviations 3) Statistical tests used. 	<p>The study did not report the following:</p> <ol style="list-style-type: none"> 4) Means 5) Standard deviations 6) Statistical tests used. 	This data is needed by the reviewer to calculate effect sizes to be included in the review.

Criterion	Criteria for inclusion	Criteria for exclusion	Rationale
6. Sample population	Participants in the study must attend a primary school/elementary school. Participants are aged between 7 and 11 years old.	Participants in the study do not attend a primary school/elementary school. Participants are not aged between 7 and 11 years old.	To ensure that the reviewer critically evaluates research focusing on the primary/elementary school-based Pyramid Club intervention.
7. Sample size	The study's sample size must include more than 1 child.	The study's sample size is 1 child only (single case research designs).	To increase the reliability of the findings covered within this review.

Figure 1: Flow chart illustrating the literature search conducted

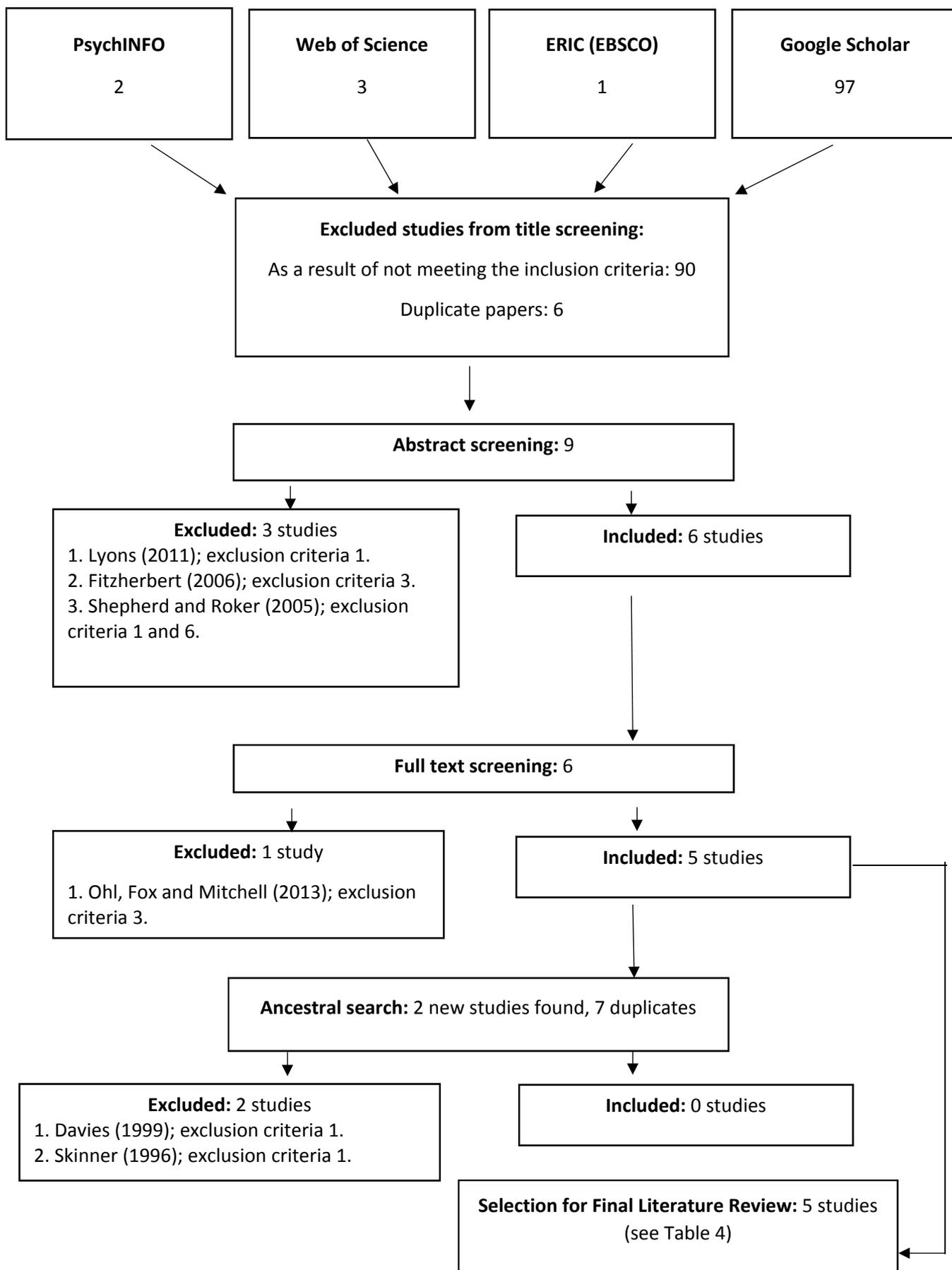


Table 4: *Studies Included Within This Review*

Study reference

-
1. Cassidy, T., McLaughlin, M. & Giles, M. (2015). Socio-emotional health in school children: an emotion-focused intervention. *International Journal of School and Cognitive Psychology*, 2 (4), 1-7.

 2. Cassidy, T., McLaughlin, M. & Giles, M. (2014). Group-based intervention to improve socio-emotional health in vulnerable children. *Journal of Psychology and Clinical Psychiatry*, 1 (7), 1-9.

 3. Ohi, M., Mitchell, K., Cassidy, T. & Fox, P. (2008). The pyramid club primary school-based intervention: evaluating the impact on children's social-emotional health. *Child and Adolescent Mental Health*, 13 (3), 115-121.

 4. McKenna, A.E., Cassidy, T. & Giles, M. (2014). Prospective evaluation of the pyramid plus psychosocial intervention for shy withdrawn children: an assessment of efficacy in 7- to 8-year-old school children in Northern Ireland. *Child and Adolescent Mental Health*, 19 (1), 9-15.

 5. Ohi, M., Fox, P. & Mitchell, K. (2013). Strengthening socio-emotional competencies in a school setting: data from the pyramid project. *British Journal of Educational Psychology*, 83, 452-466.
-

Weight of Evidence (WoE)

The reviewer evaluated each of the five studies using Gough’s (2007) weight of evidence framework. This framework allowed the reviewer to judge each of the studies in three different areas (quality of its methodology for WoE A; relevance of the methodology used for the chosen review question for WoE B; and the study’s relevance for answering the review question for WoE C). Both WoE B and WoE C were specific to this review and the reviewer selected relevant criteria to include within these weightings.

The overall value of each of the studies was calculated by averaging the scores across all of the WoE measures to form WoE D. Appendix C contains further information about the WoE calculations and descriptors used to evaluate the studies. An example of the coding protocol used for WoE A can also be found in Appendix E.

Table 5: *The Reviewer’s Weight of Evidence Judgements*

Study	Quality of methodology (WoE A)	Relevance of the methodology (WoE B)	Relevance for the review question (WoE C)	Overall weighting (WoE D)
Cassidy, McLaughlin and Giles (2015)	2 (Medium)	3 (High)	3 (High)	2.67 (High)
Cassidy, McLaughlin and Giles (2014)	2 (Medium)	3 (High)	3 (High)	2.67 (High)
Ohl, Mitchell, Cassidy and Fox (2008)	1 (Low)	2 (Medium)	3 (High)	2 (Medium)
McKenna, Cassidy and Giles (2014)	0.75 (Low)	2 (Medium)	1 (Low)	1.25 (Low)
Ohl, Fox and Mitchell (2013)	1 (Low)	2 (Medium)	3 (High)	2 (Medium)

Participants

The studies selected for review investigated the effectiveness of the Pyramid Club intervention with 1071 participants in total. All studies were conducted in the UK, three in Northern Ireland (Cassidy, McLaughlin & Giles, 2014; 2015, McKenna, Cassidy & Giles, 2014), one in West London (Ohl, Mitchell, Cassidy & Fox, 2008) and one included samples from both West London and Salford in Greater Manchester (Ohl, Fox & Mitchell, 2013). Sample sizes in these studies ranged from 385 (Ohl, Fox & Mitchell, 2013) to 82 (McKenna, Cassidy & Giles, 2014). Four studies sampled pupils aged 7-9 years. Only Cassidy, McLaughlin and Giles (2014) included a sample of 11-year-old pupils, and so findings for this age group should be treated with caution. Cassidy, McLaughlin and Giles (2015) duplicated this data for this group of participants and this was excluded from this review.

All participants were recruited from Primary or Elementary schools. No specific information was provided about the characteristics of these educational settings. Studies selected for review did not explicitly state their sampling method. However, it could be assumed that opportunity sampling was employed as participants were likely recruited from schools who were already known to be running this intervention. This limits the generalisability of the sample (Barker, Pistrang & Elliot, 2016).

The studies provided limited information about the participants' characteristics. Ohl, Fox and Mitchell's (2013) study recruited participants who attended schools situated in areas that were categorised as having 'significant need.' Therefore, a significant proportion of the children were eligible for free school meals funding (37% of the West London sample and 23% of the Salford sample). When reviewing the samples included in all of the five studies collectively, 52% of the participants were male. Data from four studies show that 43.5% of the participants who were invited to attend the

Pyramid Club intervention were male (Cassidy, McLaughlin & Giles, 2014; 2015; McKenna, Cassidy & Giles, 2014; Ohl, Fox & Mitchell, 2013). A recent meta-analytic review reported small but significant gender differences in emotion expression, with girls showing more internalising emotions and boys showing more externalising emotions (Chaplin & Aldao, 2013). This may explain why a larger proportion of girls were selected to attend the Pyramid Club in these studies, particularly as individuals with comorbid externalising behaviours were not invited to attend the intervention.

Power analyses were used to establish whether the sample sizes used in each study had enough power to detect effect sizes. The analysis conducted was based on being able to establish a medium effect size at power 0.8 with an alpha level of 0.5 (Cohen, 1992). Three studies had sufficient power (Cassidy, McLaughlin & Giles, 2014; 2015; Ohl, Fox & Mitchell, 2013), but two studies were underpowered (Ohl, Mitchell, Cassidy & Fox, 2008; McKenna, Cassidy & Giles, 2014). The findings from these two studies should therefore be considered cautiously.

Research Design

Both experimental and quasi-experimental designs formed part of this review. All studies employed group-based designs. Evidence hierarchies and typologies favour the use of Randomised Control Trials to evaluate the efficacy of interventions (Petticrew & Roberts, 2003). However, due to the selective nature of the Pyramid Club intervention, it is not possible to employ a completely randomised design as participants cannot be randomly allocated to either the intervention or control group.

The Pyramid intervention model details a clear selection process to identify children who would benefit from this intervention. This was used consistently in all studies. Two studies randomly assigned participants identified as being suitable for the

intervention to either the intervention group or a wait-list control group (Cassidy, McLaughlin & Giles, 2014; 2015).

However, the remaining three studies implemented a non-equivalent groups design and had a non-problem comparison control group made up of participants who did not require the intervention or displayed comorbid externalising behaviours. This is because spaces were available for all children who had been identified to attend a Pyramid Club, and due to ethical issues associated with assigning participants to a wait-list control group (Ohl, Mitchell, Cassidy & Fox, 2008; McKenna, Cassidy & Giles, 2014; Ohl, Fox & Mitchell, 2013). No attempts were made to establish group equivalence. This is a threat to the internal validity of the studies as the groups may have differed at baseline in terms of the participants' individual characteristics. This could have subsequently affected the findings reported (Barker, Pistrang & Elliot, 2016). These three studies received a 'medium' rating for WoE B to reflect this.

Intervention

As part of WoE C, the reviewer rated each study in relation to the quality of the description provided about the Pyramid Club implementation. All of the studies explicitly referred to the same Pyramid Club Intervention Model as outlined in the introduction, which included information about the activities that form part of each of the after-school sessions. Therefore, four out of the five studies under review received a 'high' rating.

The reviewer also evaluated the studies in relation to the fidelity of the intervention implementation. None of the studies provided information about the adults who ran the intervention. But, all of the studies except McKenna, Cassidy and Giles (2014) reported that training had been delivered to these members of staff by Pyramid Co-

ordinators prior to intervention implementation. A Pyramid Co-ordinator also visited the settings twice during the intervention period to offer ongoing supervision and ensure group leaders were following the Pyramid Club manual. Due to the lack of specificity about the fidelity of the intervention, McKenna, Cassidy and Giles' (2014) study received a 'low' rating for WoE C.

Measures

As part of evaluating studies for WoE A, the reviewer considered the reliability and validity of the measures utilised by the researchers. All studies included in the review used the SDQ as a measure to evaluate the children's social-emotional wellbeing. This measure provides a total difficulties score as well as scores for five sub-scales (emotional symptoms, conduct problems, hyperactivity, peer relationship problems, prosocial behaviour). Three out of the five studies reviewed all of these separate measures (Cassidy, McLaughlin & Giles, 2014; 2015; Ohl, Fox & Mitchell, 2013). Cassidy, McLaughlin & Giles (2014) also categorised SDQ scores into 'internalising' and 'externalising' scales. McKenna, Cassidy and Giles (2014) focused solely on three of the SDQ sub-scales (emotional symptoms, peer problems and prosocial behaviours). Only one study used the self-report version of the SDQ as the participants were aged 11 years (Cassidy, McLaughlin & Giles, 2014). All of the other studies used the teacher rating versions of the SDQ only. In addition, two studies also used the Trait Emotional Intelligence Questionnaires (TEIQue; Petrides, 2009). Cassidy, McLaughlin and Giles (2015) used the TEIQue-360S, The TEIQue-CSF was used by Cassidy, McLaughlin and Giles (2014; 2015) alongside the SDQ measure. Studies who used multiple sources and/or methods of data collection received a higher rating for WoE A.

None of the studies provided specific reliability coefficients for the SDQ. However, Goodman (1997; 1999; 2001) has reported satisfactory reliability and validity levels for this measure. Cassidy, McLaughlin and Giles (2015) reported that the Cronbach alpha statistic for the TEIQue-360S teacher rating scale used in the study was 0.95 which is considered acceptable for social science research (Field, 2013). Other researchers have reported acceptable reliability for the TEIQue-CF (Mavroveli, Petrides, Sangareau & Furnham, 2008; Mavroveli, Petrides, Shove & Whitehead 2008).

Findings

In order to address the review question, the reviewer focused only on measures which would give an indication of children's social-emotional wellbeing, as well as those which focused on internalising emotions or behaviours. As a result, Conduct problems and Hyperactivity sub-scales of the SDQ were omitted from the effect size analysis.

Table 6 displays the effect sizes calculated for each of the five studies. The Pre-Post Control Group Standardised Mean Difference (PPC SMD) measure was used to calculate these effect sizes. This measure enabled the reviewer to compare the changes reported in the outcome measures between the two groups (Morris, 2007). Each of the studies measured participant outcomes, for both intervention and control/comparison groups, at pre-intervention and post-intervention. Three of the included studies also included a follow up measure at 12 weeks (Cassidy, McLaughlin & Giles, 2014; 2015; McKenna, Cassidy & Giles, 2014).

It is important to note that there is a lack of clarity about the actual number of participants represented in the data reported in the McKenna, Cassidy and Giles

(2014) paper. The effect sizes reported are based on 88 participants, 57 in the intervention group and 31 in the comparison group.

Cohen's (1992) descriptors were used to make a judgement about the effect sizes reported (Small = 0.2; Medium = 0.5; Large = 0.8). Effect sizes reported as lower than 0.2 were described as 'not practically significant'. Table 6 provides a summary of the effect sizes for each study.

Table 6: Effect Sizes and Their Descriptors

Study	N	Source of evidence	Outcome	Pre-intervention vs. Post-intervention effect size (PPC SMD- Intervention vs. control group)	Effect size descriptor	Pre-intervention vs. Follow-up effect size (PPC SMD- Intervention vs. control group)	Effect size descriptor	WoE D
Cassidy, McLaughlin and Giles (2015)	226	Teacher ratings –SDQ	Emotional difficulties	-2.64	<i>Large</i>	-2.08	<i>Large</i>	2.67 (High)
			Peer problems	-2.47	<i>Large</i>	-2.53	<i>Large</i>	
			Prosocial behaviours	5.09	<i>Large</i>	2.86	<i>Large</i>	
			Total difficulties	-1.96	<i>Large</i>	-1.65	<i>Large</i>	
		TEIQue 360S + CSF	Emotional IQ	2.25	<i>Large</i>	1.15	<i>Large</i>	
Cassidy, McLaughlin and Giles (2014)	294	Teacher ratings – SDQ	Emotional difficulties	-0.46	<i>Small</i>	-1.20	<i>Large</i>	2.67 (High)
			Peer problems	-0.46	<i>Small</i>	-1.47	<i>Large</i>	
			Prosocial behaviours	0.22	<i>Small</i>	0.52	<i>Medium</i>	
			Total difficulties	-0.57	<i>Medium</i>	-1.04	<i>Large</i>	
			Internalising	-0.63	<i>Medium</i>	-1.54	<i>Large</i>	
			Externalising	-0.35	<i>Small</i>	-0.79	<i>Large</i>	
		Self-report ratings – SDQ	Emotional difficulties	-0.77	<i>Medium</i>	-1.24	<i>Large</i>	
			Peer problems	-0.85	<i>Large</i>	-1.18	<i>Large</i>	
			Prosocial behaviours	0.67	<i>Medium</i>	0.67	<i>Medium</i>	
			Total difficulties	-0.97	<i>Large</i>	-1.29	<i>Large</i>	
			Internalising	-1.02	<i>Large</i>	-1.33	<i>Large</i>	
			Externalising	-0.44	<i>Small</i>	-0.41	<i>Small</i>	
			TEIQue CSF	Emotional IQ	1.83	<i>Large</i>	1.23	

Study	N	Source of evidence	Outcome	Pre-intervention vs. Post-intervention effect size (PPC SMD- Intervention vs. control group)	Effect size descriptor	Pre-intervention vs. Follow-up effect size (PPC SMD- Intervention vs. control group)	Effect size descriptor	WoE D
Ohl, Mitchell, Cassidy and Fox (2008)	94	Teacher ratings – SDQ	Total difficulties	-0.70	<i>Medium</i>	--	--	2 (Medium)
McKenna, Cassidy and Giles (2014)	88	Teacher ratings – SDQ	Emotional difficulties	-0.47	<i>Small</i>	-0.57	<i>Medium</i>	1.25 (Low)
			Peer problems	-0.30	<i>Small</i>	-0.62	<i>Medium</i>	
			Prosocial behaviours	0.12	<i>Not practically significant</i>	0.29	<i>Small</i>	
Ohl, Fox and Mitchell (2013)	375	Teacher ratings - SDQ	Total difficulties	-0.35	<i>Small</i>	--	--	2 (Medium)
			Emotional difficulties	-0.66	<i>Medium</i>	--	--	
			Peer problems	-0.33	<i>Small</i>	--	--	
			Prosocial behaviours	0.22	<i>Small</i>	--	--	

Note: N represents the number of participants included in the final data analysis for each study. Effect sizes (Pretest-Posttest-Control Standardised Mean Difference; abbreviated as PPC SMD) were calculated using the following method (Morris, 2007):

- 1) Calculating the mean difference between the pre vs. post-intervention scores and pre vs. follow-up separately for both the intervention and comparison/control groups
- 2) Calculating the difference between these two differences for the intervention and comparison/control groups
- 3) The between-group difference was then divided by the pooled SD of the intervention and comparison/control group's pre-intervention scores (known as Hedge's g).

Larger positive values indicate that the intervention group's scores were higher compared to the comparison/control group, whereas larger negative values indicate that the intervention group's scores were lower compared to the comparison/control group. A score of 0 indicates that there was no difference between the mean outcome scores for the two groups.

Overall, the findings highlight a wide range of generally positive effects for the Pyramid Club intervention on the social-emotional wellbeing of primary school aged children. This intervention was found to be most effective at increasing participants' Trait Emotional Intelligence scores (Cassidy, McLaughlin & Giles, 2014; 2015; both with a 'high' WoE D), and at reducing total difficulties scores (Cassidy, McLaughlin & Giles, 2014; 2015), peer problems and emotional difficulties scores between pre-intervention and follow-up (Cassidy, McLaughlin & Giles, 2014; 2015; McKenna, Cassidy & Giles, 2014). More varied findings were reported between pre-intervention and post-intervention for reduction in emotional difficulties (effect sizes ranging from -0.46 to -2.64), peer problems (-0.30 to -2.47), total difficulties score (-0.35 to -1.96) across all of the studies.

A mixture of findings were reported for effects of the intervention on prosocial behaviours with effect sizes ranging from not practically significant to large between pre-intervention and post-intervention, and small to large between pre-intervention and follow-up. One study separately reported outcomes for the internalising and externalising scales on the SDQ (Cassidy, McLaughlin & Giles (2014). This study highlighted a more positive impact of the intervention on internalising behaviours (medium to large effect sizes) compared to externalising behaviours (small to large effect sizes). This is understandable as the intervention is aimed at children who demonstrate internalising behaviours and is regarded as unsuitable for children with comorbid externalising behaviours.

For the teacher-reported SDQ scores, larger effect sizes were reported for the reduction in both internalising and externalising problems between pre-intervention and follow-up (-1.54 and -0.79) compared to pre-intervention and post-intervention (-0.63 and -0.35). This could indicate that the teachers perceived that the intervention

group continued to make improvements in these scales after the intervention had stopped, while the internalising and externalising problems reported for the wait-list controls (who didn't receive any intervention during the data collection period) were perceived to have little to no improvement, or worsen. Cassidy, McLaughlin and Giles (2014) explained that the significant reduction reported for externalising problems could be explained by the existence of a third comorbidity category outside of pure internalising and externalising problems. They suggest that the intervention could target this comorbidity aspect too, and so it may have a positive impact on externalising problems as well as internalising problems.

Two of the studies included in the review also compared scores on Trait Emotional Intelligence between the Pyramid Club intervention group and the control group (Cassidy, McLaughlin & Giles, 2014; 2015). Large effect sizes were reported between pre-intervention and post-intervention as well as pre-intervention and follow-up on both the teacher-ratings and self-report ratings of Trait Emotional Intelligence. This is a very positive finding but further research is required before the findings can be generalised further.

Summary and conclusions

This review aimed to explore the effectiveness of the Pyramid Club intervention in improving children's social-emotional wellbeing. Overall, this intervention had a wide range of positive effects, with small to large effect sizes reported at post-intervention and follow-up for all of the outcome measures reviewed.

The largest effect sizes were reported for the reduction in emotional difficulties, peer problems and total difficulties scores at follow-up. This supports the idea that the Pyramid Club can have some positive, longer-term effects on children's social-emotional wellbeing. However, this should be interpreted cautiously as only three of the included studies included a follow-up measure. There is also emerging evidence for a positive effect of this intervention on children's trait emotional intelligence, with large effect sizes being reported at both post-intervention and follow-up. However, further evidence is required in order to support these initial findings before generalisation can be made as only two studies used this measure as part of their analysis. These studies received a 'high' WoE D rating. In addition, Cassidy, McLaughlin and Giles (2015) noted that they had amended the TEIQue-360s rating scale to meet their participants' needs, but reported an appropriate Cronbach's alpha value for this measure.

There are some methodological issues which limit the quality of this evidence in demonstrating the effectiveness of the Pyramid Club intervention. Firstly, as the majority of researchers failed to establish group equivalence, there are threats to the internal validity of these studies as other factors could have influenced the differences observed in the outcome measures between the two groups at post-intervention and follow-up. Secondly, the same researchers carried out the majority of the studies selected for this review. This is a considerable limitation of the

evidence and provides a potential source of bias as these researchers may be motivated to promote the use of this intervention. This is mediated to some extent by the use of standardised measures and involvement of school staff in providing outcome assessment data.

Although some promising positive outcomes have been reported as part of this review, these need to be considered tentatively due to the methodological limitations of the research reviewed. The research base would benefit from having further research from alternative researchers to reduce any potential bias that exists in the current research base. Future research should also aim to include multiple data collection methods where possible in order to facilitate triangulation of the findings. This could involve the exploration of the impact of this intervention on trait emotional intelligence. Group equivalence needs to be established in order for researchers to be able to draw more causal conclusions from the research. Employing an active control group would also facilitate this. Finally, longer-term follow-up assessments should be incorporated in all future research where possible to further explore the potential long-term effects of this intervention. This is particularly important for EPs and schools to consider when deciding whether to implement this type of preventative intervention approach.

Based on the evidence reviewed and the costs involved in running this intervention, it can be concluded that the Pyramid Club intervention could be a cost-effective preventative intervention to use to address maladaptive internalising behaviours in primary school aged children. These initial findings suggest that EPs should explore the use of this short-term, small group intervention further, particularly as the reported positive effects that it has on children's social and emotional wellbeing could be maintained or continue to improve after the intervention has stopped. This

could indicate that the intervention has the capacity to equip children with the skills and strategies to bring about long-term change in their social and emotional wellbeing. School-based preventative interventions targeting internalising problems are needed, particularly as these difficulties commonly go undetected and similarly to externalising problems, present a risk for mental health difficulties in later life. There are limited effective preventative interventions which target these concerns (Bayer et al., 2009). Further research evidence investigating its success with other age groups, as well as its potential positive effects on externalising problems too, would help to further inform EP Practice in this area.

References

- Barker, C., Pistrang, N., & Elliott, R. (2016). *Research Methods in Clinical Psychology: An introduction for students and practitioners* (3rd edition). Chichester: John Wiley & Sons.
- Bayer, J., Hiscock, H., Scalzo, K., Mathers, M., McDonald, M., Morris, A., Birdseye, J. & Wake, M. (2009). Systematic review of preventive interventions for children's mental health: what would work in Australian contexts? *Australian and New Zealand Journal of Psychiatry*, 43 (8), 695-710.
- Cassidy, T., McLaughlin, M. & Giles, M. (2014). Group-based intervention to improve socio-emotional health in vulnerable children. *Journal of Psychology and Clinical Psychiatry*, 1 (7), 1-9.
- Cassidy, T., McLaughlin, M. & Giles, M. (2015). Socio-emotional health in school children: an emotion-focused intervention. *International Journal of School and Cognitive Psychology*, 2 (4), 1-7.
- Chaplin, T. M., & Aldao, A. (2013). Gender Differences in Emotion Expression in Children : A Meta-Analytic Review. *Psychological Bulletin*, 139(4), 735–765.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112 (1), 155-159.
- ContinYou (2017). About Pyramid. Retrieved from <https://www.continyou.org.uk/about-pyramid/>, accessed on 4th February 2017.
- DfE & DoH (Department for Education & Department of Health) (2015). *Special educational needs and disability code of practice: 0 to 25 years*. London: The Stationery Office.
- DfE (Department for Education) (2016). *Mental health and behaviour in schools: departmental advice for school staff*. London: DfE. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/508847/Mental_Health_and_Behaviour_-_advice_for_Schools_160316.pdf, accessed on 5th February 2017.
- DH (Department of Health) (2015). *Future in Mind: Promoting, protecting and improving children and young people's mental health and wellbeing*. London: DH. Retrieved from www.gov.uk/government/uploads/system/uploads/attachment_data/file/414024/Childrens_Mental_Health.pdf, accessed on 5th February 2017.
- Dunsmuir, S. & Cobbald, A. (2017). A framework for promoting child mental health in schools. In Kelly, B., Woolfson, L.M. & Boyle, J. (Eds) *Frameworks for Practice in Educational Psychology: A textbook for trainees and practitioners* (2nd edition). (pp 309-331). London: Jessica Kingsley Publishers.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th Edition). London: Sage Publications Ltd.

- Fitzherbert, K. (1997). Promoting Inclusion: The work of the Pyramid Trust. *Emotional and Behavioural Difficulties*, 2 (3), 30-35.
- Frederickson, N. & Cline, T. (2015). *Special Educational Needs, Inclusion and Diversity (3rd edition)*. Maidenhead: Open University Press.
- 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.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, 40, 791-801.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Child and Adolescent Psychiatry*, 40(11), 1137-1145.
- Kratchowill, T.R. (2003). Task Force on Evidence-Based Interventions in School Psychology: Procedural and Coding Manual. Retrieved from http://www.indiana.edu/~ebi/documents/_workingfiles/EBImanual1.pdf, accessed on 11th February 2017.
- Lyons, R. E. (2011). *An evaluation of the use of a pyramid club to support shy and withdrawn children's transition to secondary school*. Unpublished Doctoral Dissertation, University of Manchester. Retrieved from <https://www.escholar.manchester.ac.uk/uk-ac-man-scw:128806>
- Mavroveli, S., Petrides, K.V., Sangareau, Y. & Furnham, A. (2008). Relating trait emotional intelligence to objective socioemotional outcomes in childhood. *British Journal of Educational Psychology*, 79 (2), 259-272.
- Mavroveli, S., Petrides, K.V., Shove, C. & Whitehead, A. (2008). Validation of the construct of trait emotional intelligence in children. *European Child and Adolescent Psychiatry*, 17 (8), 516-526.
- McKenna, A.E., Cassidy, T. & Giles, M. (2014) Prospective evaluation of the pyramid plus psychosocial intervention for shy withdrawn children: an assessment of efficacy in 7- to 8-year-old school children in Northern Ireland. *Child and Adolescent Mental Health*, 19 (1), 9-15.
- Morris, S. B. (2007). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364–386.
- Ohl, M., Mitchell, K., Cassidy, T. & Fox, P. (2008) The pyramid club primary school-based intervention: evaluating the impact on children's social-emotional health. *Child and Adolescent Mental Health*, 13 (3), 115-121.

Ohl, M., Fox, P. & Mitchell, K. (2013) Strengthening socio-emotional competencies in a school setting: data from the pyramid project. *British Journal of Educational Psychology*, 83, 452-466

Petrides, K. V. (2009). Technical manual for the Trait Emotional Intelligence Questionnaires (TEIQue) (1st edition, 4th printing). London: London Psychometric Laboratory.

Petticrew, M. & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health*, 57 (7), 527-529.

Pyramid (2007) *Pyramid scheme club leader training manual*. London, UK: Pyramid.

Rubin, K. H., Coplan, R.J. & Bowker, J.C. (2009). Social withdrawal in childhood. *Annual Review of Psychology*, 60, 141-171.

Seligman, M.E.P., Steen, T.A., Park, N., & Peterson, C. (2005). Positive Psychology Progress. *American Psychologist*, 60 (5), 410–421.

UWL (University of West London; 2017a) Pyramid - How it helps. Retrieved from <http://www.uwl.ac.uk/pyramid/how-it-helps>, accessed on 4th February 2017.

UWL (University of West London; 2017b) Pyramid - About. Retrieved from <http://www.uwl.ac.uk/pyramid/about>, accessed on 4th February 2017.

Appendices

Appendix A	<i>Studies excluded from the review with rationale</i>	Page 34
Appendix B	<i>Mapping of the field</i>	Pages 35-40
Appendix C	<i>Weight of evidence information</i>	Pages 41-46
Appendix D	<i>Coding protocol adaptations with rationale</i>	Page 47
Appendix E	<i>Coding protocol examples</i>	Pages 48-55

Appendix A: Studies excluded from this Literature Review with rationale

Studies excluded at abstract screening:

Study	Rationale
Lyons, R. E. (2011). <i>An evaluation of the use of a pyramid club to support shy and withdrawn children’s transition to secondary school</i> . Unpublished Doctoral Dissertation, University of Manchester. Retrieved from Google Scholar.	Exclusion criteria 1: This paper was not published in a peer-reviewed journal. It is a Doctoral Dissertation which is published on the University of Manchester’s website.
Fitzherbert, K. (2006). Promoting inclusion : the work of the pyramid trust. <i>Emotional and Behavioural Difficulties</i> ; 3 (2), pp 30-35. Retrieved from Google Scholar.	Exclusion criteria 3: The study does not contain primary data on the Pyramid Club intervention. It is a review.
Shepherd, J. & Roker, D. (2005) <i>An evaluation of a ‘transition to secondary school’ project run by the National Pyramid Trust</i> . Brighton, Trust for the Study of Adolescence. Retrieved from Google Scholar.	Exclusion criteria 1: This study was not published in a peer-reviewed journal.

Studies excluded at full text screening:

Study	Rationale
Ohl, M., Fox, P., & Mitchell, K. (2013). The Pyramid Club Elementary School-Based Intervention : Testing the Circle Time Technique to Elicit Children’s Service Satisfaction, <i>Journal of Educational and Developmental Psychology</i> , 3(2), 204–214. Retrieved from Google Scholar.	Exclusion criteria 3: The study focused on gaining pupil views of the experience of the Pyramid club, as opposed to collecting quantitative pre and post intervention data for the effectiveness of the intervention.

Studies excluded at ancestral screening:

Study	Rationale
Davies, J.H. (1999) Children’s writing improvements following participation in the Pyramid Scheme. Unpublished report, University of Surrey, Surrey.	Exclusion criteria 1: This study is an unpublished report, therefore it does not feature in a peer-reviewed journal.
Skinner, C. (1996) Evaluation of the effectiveness of Pyramid Clubs held in 1995-6. Unpublished report, University of Surrey.	Exclusion criteria 1: This study is also an unpublished report. So it does not meet the criteria required to be included within the review as it has not been published in a peer-reviewed journal.

Appendix B – Table to Summarise the Included Studies Used In This Review (Mapping the Field)

Study authors	Country	Sample characteristics	Study design	Intervention information	Outcome measures	Key findings related to outcome measures relevant to this review
Cassidy, McLaughlin and Giles (2015)	UK – Northern Ireland	<p>(a) *Focus for this review* 226 Year 4 children; 101 boys and 125 girls. Intervention group: 141 (75 girls, 66 boys). Wait list controls: 85 (50 girls and 35 boys).</p> <p>(b) *Not included in the review* 294 Year 7 children; 122 boys and 172 girls. Intervention group: 162 (100 girls and 62 boys). Wait list controls: 132 (72 girls and 60 boys).</p>	<p>Intervention group vs. wait-list control group (random allocation).</p> <p>Pre vs. post-intervention vs. 12 week follow up measures.</p>	<p><u>Pyramid Club</u></p> <ul style="list-style-type: none"> - Use of the 3 stage model to screen/identify children suitable for the intervention - Manualised intervention (Pyramid, 2007) - Implemented 10 weekly sessions, 90 minutes long implemented during the school day. - Ongoing supervision by Pyramid Club Coordinator. - Training provided to club leaders prior to intervention. - Visits to clubs twice in the 10 week period. 	<p>Strengths and Difficulties Questionnaire (Goodman, 1997; 1999)-</p> <ul style="list-style-type: none"> - Pre- intervention, post-intervention and follow-up Teacher rating questionnaire used for Year 4 participants. - Teacher and self-ratings used for Year 7 pupils. Scores provided for each of the SDQ sub-scales, as well as the total difficulties score <p>The Trait Emotional Intelligence Questionnaire Child-Teacher rating version (TEIQue-360S; Petrides, 2009).</p> <ul style="list-style-type: none"> - Measures 15 aspects of emotional intelligence. - Raters read a statement for each aspect of emotional intelligence and rate the child accordingly. - The rating scale was adapted in this study to meet the needs of the participants (a 5 point Likert scale was used). <p>The Trait Emotional Intelligence Questionnaire Child Short Form (TEIQue-CSF; Petrides, 2009).</p> <ul style="list-style-type: none"> - Measures 9 of the 15 aspects of emotional intelligence. - Completed by the child. - Pre- intervention, post-intervention and follow-up 	<p>Year 4 participants- <u>SDQ outcomes-</u></p> <p>(a) Significant main effects and interaction effects at $p < 0.001$ for total difficulties, emotional problems, peer problems, pro-social behaviour. Post hoc analyses revealed that these effects were for the Pyramid group members only between time 1 and 2 and time 1 and 3.</p> <p>(b) Large effect sizes reported at post intervention and follow-up.</p> <p><u>Trait Emotional intelligence outcomes-</u></p> <p>(a) Main effects were reported between the Pyramid and Wait-List groups at post-intervention ($p < 0.001$) and follow up ($p < 0.001$).</p>

Study authors	Country	Sample characteristics	Study design	Intervention information	Outcome measures	Key findings related to outcome measures relevant to this review
Cassidy, McLaughlin and Giles (2014)	UK – Northern Ireland	294 pupils in year 7 (aged 11) attending 13 schools who participated in the study. 122 boys, 172 girls. Intervention group: 162, (100 girls, 62 boys) Wait list controls: 132 (72 girls, 60 boys). Data was collected over two academic years.	Intervention group comparison with a waiting list control group (random allocation). Pre vs. post-intervention vs. 12 week follow up measures.	<p><u>Pyramid Club</u></p> <ul style="list-style-type: none"> - Use of the 3-stage model to screen/identify children suitable for the intervention in line with the Pyramid intervention manual (use of SDQ scores and professional knowledge of risk factors for internalisation associated with the children’s lives). - Manualised intervention (Pyramid, 2007) - Implemented 10 weekly sessions, 90 minutes long implemented during the school day. - Ongoing supervision by Pyramid Club Coordinator. - Training provided to club leaders prior to intervention. - Visits to clubs twice in the 10 week period. 	<p>Strengths and Difficulties Questionnaire (Goodman, 1997;1999)-</p> <ul style="list-style-type: none"> - Pre- intervention - Post-intervention - Follow up <p>Teacher and self-ratings used.</p> <p>Scores provided for each of the SDQ sub-scales, including internalising and externalising scales, as well as the total difficulties score.</p> <p>The Trait Emotional Intelligence Questionnaire Child Short Form (TEIQue-CSF; Petrides, 2009).</p> <ul style="list-style-type: none"> - Pre- intervention - Post-intervention - Follow up 	<p><u>SDQ outcomes-</u></p> <p>Post hoc comparisons revealed effects reported in the ANOVA were only seen for the Pyramid Club group. Significant differences in time were between time 1 and both time 2 and time 3.</p> <ul style="list-style-type: none"> (a) Pyramid Club produced significant reductions in total difficulties scores and prosocial behaviours for self-rated and teacher rated scales. This was maintained at the follow-up analysis. (b) Effects were mainly found for emotional difficulties and peer difficulties for both teacher and self-ratings. Interaction effects were also reported for internalising behaviours. (c) Small interaction effect on self and teacher ratings for externalising problems. (d) Large effect size (teacher and self-ratings) observed for emotional problems, peer problems, pro-social behaviours and internalising problems. (e) A significant effect on emotional intelligence scores were found – scores increased between time 1 and time 2, and were maintained at the follow-up analysis.

Study authors	Country	Sample characteristics	Study design	Intervention information	Outcome measures	Key findings related to outcome measures relevant to this review
Ohl, Mitchell, Cassidy and Fox (2008)	UK – West London	<p>105 children (43 Pyramid Club participants and 62 non-problem comparison participants).</p> <p>51 girls, 54 boys aged 8-9 years in Year 3 classes who attended four West London Primary Schools.</p> <p>11 Pyramid club children and 10 non-problem comparison participants had incomplete data, so complete data was analysed for 94 participants (42 children in the intervention group, and 52 children in the non-problem comparison group).</p>	<p>Intervention group comparison with a non-problem group of participants</p> <p>Pre vs. post-intervention</p>	<p><u>Pyramid Club</u></p> <ul style="list-style-type: none"> - Use of the 3-stage model to screen/identify children suitable for the intervention in line with the Pyramid intervention manual (use of SDQ scores and staff/professionals knowledge of the children). - Manualised intervention (Pyramid, 2007) - Implemented 10 weekly sessions, 90 minutes long. - Implemented by trained volunteer leaders - Ongoing supervision by Pyramid Club Coordinator. - Training provided to club leaders prior to intervention. - Visits to clubs twice in the 10 week period. <p>The children’s progress in the club was discussed at a Post-Pyramid Club multi-agency intervention meeting.</p>	<p>Strengths and Difficulties Questionnaire (Goodman, 1997;1999)-</p> <ul style="list-style-type: none"> - Pre- intervention - Post-intervention <p>Teacher ratings used.</p> <p>Only the total difficulties score was used.</p>	<p><u>SDQ outcomes-</u></p> <ul style="list-style-type: none"> (a) Both groups had a decrease in mean scores on the total difficulties scale at post-intervention. (b) Changes in total difficulties scores for the Pyramid Club children were significantly different to changes reported for the comparison group children. (c) Pyramid Club children = large effect size for decrease in outcome measures over time which was more significant compared to the comparison group (r=.71). The comparison group only had a moderate effect size (r=.44). (d) A significant main effect of time AND group was reported. <p><u>Impact of the Pyramid Club SDQ banding-</u></p> <ul style="list-style-type: none"> (a) 20 children showed an improvement in terms of their post-intervention SDQ banding, 20 remained the same and 2 worsened and entered a lower banding. This is compared to 44 children staying in the same band, 4 children improving and 4 children deteriorating in the non-problem comparison group. (b) Post-intervention shifts in SDQ banding were more reflective of the SDQ community norms compared to baseline SDQ bands.

Study authors	Country	Sample characteristics	Study design	Intervention information	Outcome measures	Key findings related to outcome measures relevant to this review
McKenna, Cassidy and Giles (2014)	UK – Northern Ireland	<p>208 Primary 4 children aged 7-8 years; (45.2% male) were screened for 2009-10 and 2010-11 intervention attendance. 126 cases were excluded due to insufficient data collection.</p> <p>Analysed sample = 82 children aged 7-8 years in 7 schools.</p> <p>Pyramid Plus intervention group = 57 children (41.7% males) Comparison group = 31 children (50.6% male).</p> <p>Comparison/control group = classmates who were screened but did not require the intervention.</p>	<p>Intervention group vs. comparison/control group</p> <p>Pre vs. post-intervention (10 week) vs. follow up (12 weeks).</p>	<p><u>Pyramid Club</u></p> <ul style="list-style-type: none"> - Use of the 3-stage model to screen/identify children suitable for the intervention in line with the Pyramid intervention manual (use of SDQ scores, with professional knowledge of risk factors for internalising behaviour problems). - Manualised intervention (Pyramid, 2007) - Implemented 10 weekly sessions, 90 minutes long. - 1:4 staff ratio, 10-12 children attended clubs. 	<p>Strengths and Difficulties Questionnaire (Goodman, 1997;1999)-</p> <ul style="list-style-type: none"> - Pre- intervention - Post-intervention - Follow-up <p>Teacher ratings used.</p> <p>Only the emotional difficulties, peer problems and prosocial behaviours sub-scales were reviewed.</p>	<p><u>SDQ outcomes-</u></p> <ul style="list-style-type: none"> (a) Changes in emotional symptoms and peer problems were found to be dependent on allocation to the intervention or control group. (b) No significant interaction effects reported for pro-social behaviour. <p><u>Scoring band shifts for individual participants:</u></p> <ul style="list-style-type: none"> (a) Amount of children in the borderline-abnormal category for emotional symptoms decreased from 33% (pre) to 6.3% (post) and 10% (follow-up) (b) Amount of children in the borderline-abnormal category for peer problems decreased from 22.8% (pre) to 3.2% (post) and 5.8% (follow-up) (c) Amount of experiencing peer exclusion decreased from 35.6% (pre) to 13.7% (post) and 24.3% (follow-up)

Study authors	Country	Sample characteristics	Study design	Intervention information	Outcome measures	Key findings related to outcome measures relevant to this review
Ohl, Fox and Mitchell (2013)	UK – 7 schools across 2 cities (Ealing and Salford).	<p>385 children aged 7-8 years old in Year 3 at Primary School. 200 boys, 185 girls.</p> <p>Intervention group: 103 children - 57 girls, 46 boys. Non-equivalent comparison group- 282 children</p> <p>Eligibility for FSM – 37% West London and 23% Salford.</p> <p>Data collected across two academic years.</p> <p>Final sample analysed - 102 Pyramid Club attendees 273 comparison group children</p>	<p>Intervention group vs. comparison/control group</p> <p>Pre vs. post-intervention (12 weeks)</p>	<p><u>Pyramid Club</u></p> <ul style="list-style-type: none"> - Use of the 3-stage model to screen/identify children suitable for the intervention in line with the Pyramid intervention manual. - Manualised intervention (Pyramid, 2007) - Implemented 10 weekly sessions, 90 minutes long. - Ongoing supervision from a Pyramid Co-ordinator. They also visited clubs twice during the 10 week implementation period to ensure fidelity. <p>Post-Pyramid Club intervention multi-agency meeting took place after the 10 weeks had elapsed to review the children’s progress. A minimum of 70% club attendance was set in order to be included within the final sample analysed.</p>	<p>Strengths and Difficulties Questionnaire (Goodman, 1997;1999)-</p> <ul style="list-style-type: none"> - Pre- intervention (baseline measure) - Post-intervention (12 weeks) <p>Teacher ratings used.</p> <p>Scores provided for each of the SDQ sub-scales, as well as the total difficulties score.</p>	<p><u>SDQ outcomes-</u></p> <p>(a) Total difficulties - A more significant decrease over time for Pyramid Club children ($p < 0.001$) = moderate effect size found ($r = .35$). Slight but non-significant increase found for the comparison children.</p> <p><u>Emotional symptoms-</u> main effect found over time ($p < 0.001$) and interaction effect ($p < 0.001$) for the Pyramid attendee group with a moderate effect size ($r = .40$). Slight but non-significant increase found for the comparison children.</p> <p><u>Peer problems-</u> No significant main effect over time but there was a significant interaction effect. Pyramid Club children’s scores decreased ($p < 0.01$, moderate effect size $r = .30$), comparison group scores increased. ANCOVA analysis however was only approaching significance ($p = 0.08$).</p> <p><u>Prosocial behaviour –</u> significant main effect over time ($p < 0.05$) and significant interaction effects ($p = 0.05$), with Pyramid Club children’s scores increasing more significantly ($p < 0.05$, effect size was modest ($r = .24$)). Comparison group had a slight, non-significant decrease. Interaction effect was more significant ($p < 0.01$) in the ANCOVA analysis.</p>

Study authors	Country	Sample characteristics	Study design	Intervention information	Outcome measures	Key findings related to outcome measures relevant to this review
						<p><u>Scoring band shifts for the SDQ:</u> (a) Post-intervention = Abnormal to normal (39% movement for the Pyramid club group) Abnormal to borderline (13% movement for the Pyramid club group). (b) Post-intervention = Abnormal to borderline (1.8% movement for the comparison group) 1 child moved from the normal band to borderline. Abnormal to borderline (13% movement for the Pyramid club group).</p>

Appendix C: Weight of Evidence information

Weight of Evidence A (WoE A) – Methodological quality

WoE A rated each of the selected studies on the quality of its research methodology. As all studies selected for the review were group-based designs, the Kratchowill (2003) coding protocol for group-based designs was used to give the ratings detailed below. The reviewer amended the coding protocol to suit the purpose of the review question (please refer to Appendix D).

Four key elements of each study were reviewed and subsequently given a numerical rating from 0-3 in relation to the quality of its methodology (3 = strong evidence; 2= promising evidence; 1 = weak evidence and 0= no evidence). To calculate an overall value for the methodological quality of the study, the scores across the four key elements were averaged.

The study must have received an average rating of 2.5 or above to receive a ‘high’ rating.

The study must have received an average rating between 1.5 or 2.4 to receive a ‘medium’ rating.

The study must have received an average rating which was 1.4 or below to receive a ‘low’ rating

Table Showing Weight of Evidence A (WoE A) Criteria (summarised from Kratchowill, 2003)

Measurement			
High (3)	Medium (2)	Low (1)	Inadequate (0)
<ul style="list-style-type: none"> • Most of the measures used have a reliability coefficient of 0.85 or more. • The researchers used multiple methods/sources for data collection. • The researchers present a case for validity for all of the measures they used. • The above criteria should be met for all of the outcome measures included in the study. 	<ul style="list-style-type: none"> • Most of the measures used have a reliability coefficient of 0.70 or more. • The researchers used multiple methods AND/OR sources for data collection. • The above criteria should be met for 75% of the outcome measures included in the study. 	<ul style="list-style-type: none"> • Most of the measures used have a reliability coefficient of 0.50 or more. • The researchers used multiple methods OR sources for data collection, but this is not necessary. • The above criteria should be met for 50% of the outcome measures included in the study. 	<ul style="list-style-type: none"> • Measures used have poor reliability scores. • The researchers used multiple methods OR sources for data collection, but this is not necessary.

Comparison group			
High (3)	Medium (2)	Low (1)	Inadequate (0)
<ul style="list-style-type: none"> • The study had an active control group. • The researchers established group equivalence. This may have been through random assignment. • Evidence of counterbalancing. • Low attrition rates were reported in the study. 	<ul style="list-style-type: none"> • The study had at least a no-intervention/wait-list control group. • 2 of the following are required: <ul style="list-style-type: none"> ○ Group equivalence ○ Low attrition rates (or an analysis to correct for this) ○ Counterbalancing 	<ul style="list-style-type: none"> • The study included a comparison group. • 1 of the following are required: <ul style="list-style-type: none"> ○ Group equivalence ○ Low attrition rates (or an analysis to correct for this) ○ Counterbalancing 	<p>The researchers made no efforts to establish group equivalence.</p>

Implementation Fidelity			
High (3)	Medium (2)	Low (1)	Inadequate (0)
<ul style="list-style-type: none"> • Two of the following were implemented in the study: <ul style="list-style-type: none"> ○ Ongoing supervision and consultation ○ Coding of sessions ○ Audio/video recording • Use of a manual, which contains detailed information about the procedures and the sequence of activities (lesson by lesson if appropriate) OR formal training was delivered for the above. • A description for how the intervention is adapted for different contexts is included. 	<ul style="list-style-type: none"> • One of the following were implemented in the study: <ul style="list-style-type: none"> ○ Ongoing supervision and consultation ○ Coding of sessions ○ Audio/video recording • Use of a manual, which contains information detailing a broad overview of the key principles of the intervention and description of the different intervention phases OR formal training was delivered for the above. 	<ul style="list-style-type: none"> • One of the following were implemented in the study: <ul style="list-style-type: none"> ○ Ongoing supervision and consultation ○ Coding of sessions ○ Audio/video recording OR use of a manual. 	<ul style="list-style-type: none"> • Little/no evidence of the implementation of suitable measures to ensure implementation fidelity.

Follow-up			
High (3)	Medium (2)	Low (1)	Inadequate (0)
<ul style="list-style-type: none"> • The researchers conducted multiple follow-up assessments. • All participants included in the original sample were included in the follow-up assessments. • The researchers used similar outcome measures. 	<ul style="list-style-type: none"> • The researchers conducted at least one follow-up assessment. • Most of the participants included in the original sample were included in the follow-up assessments. • The researchers used similar outcome measures. 	<ul style="list-style-type: none"> • The researchers conducted at least one follow-up assessment. • Some of the participants included in the original sample were included in the follow-up assessments. 	<ul style="list-style-type: none"> • No follow up assessment was included as part of the study.

	Measurement	Comparison Group	Implementation Fidelity	Follow-up assessment	Overall value for WoE A
Cassidy, McLaughlin and Giles (2015)	2	2	2	2	2 (Medium)
Cassidy, McLaughlin and Giles (2014)	2	2	2	2	2 (Medium)
Ohl, Mitchell, Cassidy and Fox (2008)	1	1	2	0	1 (Low)
McKenna, Cassidy and Giles (2014)	1	0	1	1	0.75 (Low)
Ohl, Fox and Mitchell (2013)	1	1	2	0	1 (Low)

Summary of WoE A judgements

Weight of Evidence B (WoE B) – Relevance of the Research methodology

WoE B was used to consider the research design of each study and its relevance for evaluating the effectiveness of the Pyramid Club intervention with children in Primary or Elementary Schools. Petticrew and Roberts (2003) informed the criteria and descriptors used for this weighting. The reviewer added in a criterion related to the use of follow-up measures to WoE B as it was considered important to look at the potential long-term effects of the Pyramid club intervention in order to truly evaluate its effectiveness. This is reflected in the judgements made, with studies who included a follow-up measure being given a higher rating. All of the criteria listed below had to be met in order to receive that rating.

To receive a rating of ‘High’ (3) studies must:

- Have randomly assigned participants to the intervention or control group OR established group equivalence using statistical analyses
- Have at least one active, no treatment (wait-list control) or alternative treatment control group
- Have completed pre-intervention, post-intervention and follow-up assessments using a standardised measure

To receive a rating of ‘Medium’ (2) studies must:

- Have at least one comparison group which may not be matched to the intervention group
- Have used non-random assignment to allocate participants to the intervention and comparison group
- Have completed pre-intervention and either post-intervention or follow-up assessments using a standardised measure

To receive a rating of ‘Low’ (1) studies must:

- Have no comparison group
- Have used non-random assignment to allocate participants to the intervention group
- Have completed pre-intervention and post-intervention or follow up assessments using any type of measure

Summary of WoE B judgements

Study authors	WoE B rating
Cassidy, McLaughlin and Giles (2015)	3 (High)
Cassidy, McLaughlin and Giles (2014)	3 (High)
Ohl, Mitchell, Cassidy and Fox (2008)	2 (Medium)
McKenna, Cassidy and Giles (2014)	2 (Medium)
Ohl, Fox and Mitchell (2013)	2 (Medium)

Weight of Evidence C (WoE C) – Relevance of the Research methodology

WoE C was used to judge each of the selected studies in relation to their relevance and appropriateness for answering the review question. In order to make this judgement the reviewer carefully considered 1) the amount of training provided for Pyramid Club group leaders; 2) the fidelity of programme implementation; 3) the description of the intervention and selection of participants provided.

All of the criteria listed below each of the weightings had to be met in order to receive that rating.

To receive a rating of ‘High’ (3) studies must:

- Have stated that staff who delivered the intervention had received training from Pyramid Co-ordinators
- Have used the Pyramid manual to guide intervention implementation and selection of participants
- Have provided a clear description of the intervention programme that was delivered, including examples of the activities that made up each session
- Have reported at least one visit from a Pyramid Co-ordinator during the intervention implementation period

To receive a rating of ‘Medium’ (2) studies must:

- Have stated that staff who delivered the intervention had received training from Pyramid Co-ordinators
- Have used the Pyramid manual to guide intervention implementation and selection of participants
- Have provided a sufficient description of the intervention programme that was delivered
- Have described the level of fidelity to the Pyramid Club intervention

To receive a rating of ‘Low’ (1) studies must:

- Have used the Pyramid manual to guide intervention implementation and selection of participants
- Have provided a sufficient description of the intervention programme that was delivered
- Have not described the level of fidelity to the Pyramid Club intervention sufficiently

Summary of WoE C judgements

Study authors	WoE C rating
Cassidy, McLaughlin and Giles (2015)	3 (High)
Cassidy, McLaughlin and Giles (2014)	3 (High)
Ohl, Mitchell, Cassidy and Fox (2008)	3 (High)
McKenna, Cassidy and Giles (2014)	1 (Low)
Ohl, Fox and Mitchell (2013)	3 (High)

Weight of Evidence D – Total weighting of the study

Scores across WoE A, B and C were averaged in order to calculate WoE D. For an overall rating of ‘high’, WoE D must be 2.5 or above. For an overall rating of ‘medium’, WoE D must be between 1.5 and 2.4. For an overall rating of ‘low’, WoE D must be 1.4 or below.

Summary of WoE D judgements

	Quality of methodology (WoE A)	Relevance of the methodology (WoE B)	Relevance for the review question (WoE C)	Overall weighting (WoE D)
Cassidy, McLaughlin and Giles (2015)	2 (Medium)	3 (High)	3 (High)	2.67 (High)
Cassidy, McLaughlin and Giles (2014)	2 (Medium)	3 (High)	3 (High)	2.67 (High)
Ohl, Mitchell, Cassidy and Fox (2008)	1 (Low)	2 (Medium)	3 (High)	2 (Medium)
McKenna, Cassidy and Giles (2014)	0.75 (Low)	2 (Medium)	1 (Low)	1.25 (Low)
Ohl, Fox and Mitchell (2013)	1 (Low)	2 (Medium)	3 (High)	2 (Medium)

Appendix D – Coding protocol adaptations with rationale

Adaptations made by the reviewer to the Kratchowill (2003) coding protocol, with rationale.

Section	Amendment details	Reviewer’s rationale
1 (B3 - counterbalancing)	Answer category added – not reported	This option was added so that the reviewer could code the studies selected for review for this item.
1 (B7 and B8 – qualitative research methods)	Removed	None of the studies selected for the review used qualitative data collection methods.
2 (C1-C5 – primary/secondary outcomes)	Removed	Primary/secondary outcomes and their statistical significance were reviewed later in the systematic review.
2 (D1-D4 – educational or clinical significance)	Removed	These criteria are included elsewhere in the review.
2 (E1-E7 – identifiable components)	Removed	This part of the coding protocol relates to a part which was already removed (C1-C5).
2 (G1-G3 - replication)	Removed	The reviewer judged this section to be irrelevant for the purpose of the current review.
2 (H1-H2 – site of implementation)	Removed	As all studies involved the implementation of the intervention within schools this did not need to be included within the coding protocol.
III (A2-A5 – participant characteristics)	Removed	The reviewer collated this information at a previous stage when mapping the field.
III (B- length of the intervention, C – intervention dosage, D – dosage response, E- programme implementer, F – intervener characteristics, G – intervention style and H- cost analysis data)	Removed	B – all of the studies implemented the intervention for 10-weeks, as stated in the manual. C and D were therefore considered irrelevant for the review. E and F were considered as part of WoE C. G – the same intervention was reviewed so this criteria was considered irrelevant for this review. H – was not relevant for this review.
III (J1-3 – intervention feasibility)	Removed	This was considered to be irrelevant for this review question.

Appendix E –Coding protocol example

Coding Protocol: Group-Based Design

Domain:

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

Academic intervention programs

Family and parent intervention programs

School-wide and classroom-based programs

Comprehensive and coordinated school health services

Name of Coder(s): Date: 12th February 2017

Full Study Reference in APA format: Cassidy, T., McLaughlin, M. & Giles, M. (2014) Group-based intervention to improve socio-emotional health in vulnerable children. *Journal of Psychology and Clinical Psychiatry*, 1 (7), 1-9.

Intervention Name (description from study): Pyramid Plus intervention

Study ID Number (Unique Identifier): 2

Type of Publication: (Check one)

Book/Monograph

Journal article

Book chapter

Other (specify):

—

Alpha level: 0.05

ES: Medium

N required: 86 total sample size

B4. Total size of sample (start of the study): 294

N

B5. Intervention group sample size: 162

N

B6. Control group sample size: 132

N

C. Type of Program (select one)

C1. Universal prevention program

C2. Selective prevention program

C3. Targeted prevention program

C4. Intervention/Treatment

C5. Unknown

D. Stage of the Program (select one)

D1. Model/demonstration programs

D2. Early stage programs

D3. Established/institutionalized programs

D4. Unknown

E. Concurrent or Historical Intervention Exposure (select one)

E1. Current exposure

E2. Prior exposure

E3. Unknown

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

(3=Strong Evidence; 2=Promising Evidence; 1=Weak Evidence; 0=No Evidence)

A. Measurement (answer A1 through A4)

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

A1.1 Yes

A1.2 No

A1.3 Unknown/unable to code

A2. Multi-method (select one of the following)

A2.1 Yes

A2.2 No

A2.3 N/A

A2.4 Unknown/unable to code

A3. Multi-source (select one of the following)

A3.1 Yes

A3.2 No

A3.3 N/A

A3.4 Unknown/unable to code

A4. Validity of measures reported (select one of the following)

A4.1 Yes validated with specific target group

A4.2 In part, validated for general population only

A4.3 No

A4.4 Unknown/unable to code

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

3 2 1 0

B. Comparison Group

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

B1.1 Typical contact

B1.2 Typical contact (other) specify:

B1.3 Attention placebo

B1.4 Intervention elements placebo

- B1.5 Alternative intervention
- B1.6 Pharmacotherapy
- B1.7 No intervention
- B1.8 Wait list/delayed intervention
- B1.9 Minimal contact
- B1.10 Unable to identify comparison group

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

3 2 1 0

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

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

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

- B3.1 By change agent
- B3.2 Statistical
- B3.3 Other

B4.4 Not reported/none

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

- B4.1 Random assignment
- B4.2 Posthoc matched set
- B4.3 Statistical matching
- B4.4 Post hoc test for group equivalence
- B4.5 Not reported

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

- B5.1 Low Attrition (less than 20% for Post)
- B5.2 Low Attrition (less than 30% for follow-up)

B5.3 Intent to intervene analysis carried out

No attrition rates stated; all participants included in pre, post and follow up measures.

F. Implementation Fidelity

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

F1.1 Ongoing supervision/consultation

F1.2 Coding intervention sessions/lessons or procedures

F1.3 Audio/video tape implementation (select F1.3.1 or F1.3.2):

F1.3.1 Entire intervention

F1.3.2 Part of intervention

F2. Manualization (select all that apply)

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

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

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

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

F3. Adaptation procedures are specified (select
one)

yes no unknown

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

I. Follow Up Assessment

Timing of follow up assessment: specify: 12 weeks

Number of participants included in the follow up assessment: specify: 294

Consistency of assessment method used: specify: SDQ used for pre, post and follow-up.

Rating for Follow Up Assessment (select 0, 1, 2, or 3): 3 2 1 0

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

A1. Sampling procedures described in detail yes no

Specify rationale for selection: Children were selected to be appropriate for the intervention group after reviewing their SDQ scores and discussions at a multi-agency meeting. Children with emotional difficulties and peer problems, lower incidences of prosocial behaviour demonstration and no sign of externalising behaviours were selected.

Specify rationale for sample size: _____

- A1.1 Inclusion/exclusion criteria specified yes no
- A1. Inclusion/exclusion criteria similar to school practice yes no
- A1.3 Specified criteria related to concern yes no

Training and Support Resources (select all that apply)

- 11. Simple orientation given to change agents
- 12. Training workshops conducted
 - a. of Workshops provided: Unknown
 - Average length of training: Not stated.
 - Who conducted training (select all that apply)
 - 12.1 Project Director
 - 12.2 Graduate/project assistants
 - 12.3 Other (please specify): Pyramid Club Coordinators
 - 12.3 Unknown
- 13. Ongoing technical support
- 14. Program materials obtained
- 15. Special Facilities
- 16. Other (specify):
—

Summary of Evidence for Group-Based Design Studies

Indicator	Overall evidence rating NNR = no numerical rating Or 0-3	Description of Evidence Strong Promising Weak No/limited evidence OR Descriptive ratings
General Characteristics		
General Design Characteristics	NNR	Use of a nonrandomised block design (between-participants). Very high confidence in understanding of how participants were assigned.
Statistical Treatment	NNR	Sample size sufficient to obtain a large or medium effect size at an alpha level of 0.05.
Type of Program	NNR	A selective prevention program was used.
Stage of Program	NNR	Model/demonstration
Concurrent/Historical Intervention Exposure	NNR	Unknown/unspecified
Key Features		
Measurement	2	Promising evidence - two sources of data collection (teacher rating and self-report rating for the SDQ and the TEIQue CSF). Satisfactory reliability scores for the TEIQue have been reported in previous research. Satisfactory reliability scores for the SDQ have been reported (Goodman, 1997; 1999; 2001).
Comparison Group	2	A wait-list control group was used. Children identified as being suitable for the Pyramid Club intervention were randomly allocated to either the intervention or wait-list control group.
Implementation Fidelity	2	Promising evidence – Pyramid Club Coordinators visited the Pyramid Club Leaders twice during the intervention period to check the fidelity of the programme implementation and provided ongoing supervision. The Club leaders used the Pyramid Club manual and this was checked by the Pyramid Club Coordinator.
Follow Up Assessment Conducted	2	Promising evidence – one follow up assessment was conducted. No attrition data was reported so all of the original sample appeared in the follow up analysis too. The same measures were used at follow up.