

***Case Study 1: An Evidence-Based Practice Review Report
Theme: School (setting) based interventions for children with special
educational needs (SEN)***

How effective is the 'Pyramid Club' intervention at reducing internalising behaviours and increasing prosocial skills in children aged 7-14 with identified social-emotional issues?

Summary

Pyramid Club is an intervention aimed at promoting socio-emotional well-being in children and young people (CYP) aged between 7 and 14. Children are screened for the intervention using the Strengths and Difficulties Questionnaire (SDQ; Goodman et al., 1997). CYP who are displaying 'abnormal' (clinical diagnosis) or 'borderline' (at risk) levels of internalising behaviours are deemed suitable to benefit from the intervention. The objectives of Pyramid Club map onto 3 subscales of the SDQ, with the aim to reduce 'emotional symptoms' and 'peer problems' and increase 'prosocial skills'. Pyramid Club is a low cost, UK based intervention that, if found to be effective, would be a useful resource for Educational Psychologists (EP) when supporting schools, CYP and families. Therefore, the present review aimed to synthesise the literature on Pyramid Club.

The current review identified and evaluated five studies. The findings revealed Pyramid Club is effective at reducing internalising behaviours, with medium-large effect sizes. Evidence for the improvement in prosocial skills was less conclusive, ranging from no significant improvements, to large effect sizes. These findings were discussed in the context of each study's methodological quality, relevance and limitations. Future research avenues were discussed to improve the evidence base for Pyramid Club.

Introduction

Rationale and Relevance

The Special Educational Needs and Disability (SEND) Code of Practice (CoP; Department for Education & Department of Health, 2015) outlines four broad areas of SEND, one of which is Social, Emotional and Mental Health (SEMH) difficulties. Within this definition, the CoP considers a child with an SEMH need may become “withdrawn or isolated, as well as displaying challenging, disruptive or disturbing behaviour. These behaviours may reflect underlying mental health difficulties” (Department for Education & Department of Health, 2015, p98, para 6.32).

The association between poor childhood socio-emotional functioning and later life mental health disorders has been established in a number of studies (e.g. Thomson et al., 2019; Goodwin et al., 2004). In childhood, mental health issues most commonly manifest as internalising problems (e.g. shyness and withdrawal) and externalising problems (e.g. aggression and defiance; Hiscock et al., 2012). Given these findings and in the context of Covid-19, an intervention that targets dimensions of these behaviours which is low cost, can be implemented by schools ‘in house’ and is proven to be effective, would provide a vital tool to Educational Psychologists (EP) and the schools, CYP and families they support.

Pyramid Club

Pyramid Club is an intervention designed to promote socio-emotional well-being in children and young people (CYP) aged between 7 and 14 years old. The intervention targets children who may be quiet and shy, who are more likely to internalise problems. Pyramid Club runs once a week as a 10-week school-based club, with 90-minute

sessions to aid children’s coping skills, emotional resilience, confidence and friendship skills. Main elements of the intervention are detailed in Table 1 (adapted from Ohl et al., 2008 and Jayman et al., 2019a). Up to 12 children can attend a club, with a suggested adult to child ratio of 1:3. Children are screened for the intervention on a whole class level, a multi-disciplinary meeting takes place to discuss and allocate children who are appropriate for the intervention, clubs commence for 10 weeks and then a final multi-disciplinary meeting takes place to discuss the children’s progress. Club leaders can be trained volunteers, paid staff or existing school staff.

Table 1

Main Elements of the Intervention

| Week | Therapeutic Activity | Aim |
|------|----------------------------------|---|
| 1 | Naming and creating group rules | Encourages a sense of belonging and ownership of the club |
| 1-10 | Circle time | Facilitates talking, listening, turn-taking and encourages expression of feelings |
| | Arts/crafts | Facilitates self-expression and sense of achievement |
| | Games/physical activity | Practice social and cooperation skills |
| | Food/snack | Opportunity to learn about sharing, social skills and nurturing. |
| 10 | Closing the group and reflection | A celebration of achievement in the club saying ‘thank you’ and ‘goodbye’ |

Children are screened for the intervention using the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; see Appendix 1 for more information). Scores

on the SDQ have been found to be correlated to scores on mental health measures (Goodman & Goodman, 2009) and to be predictive of mental health difficulties, with a sensitivity of 74.6% for depressive disorders and 50.5% for anxiety disorders (Goodman et al., 2000). Bryant et al. (2020) suggest the SDQ is a valuable screening measure for identifying mental health issues in children who are struggling. The aims of Pyramid Club map onto three subscales of the SDQ ('emotional symptoms', 'peer problems' and 'prosocial skills').

Children who score within the SDQ categories of 'abnormal' (clinical diagnosis) and 'borderline' (at risk of clinical diagnosis) for 'internalising behaviours' are considered appropriate for the intervention. If there are sufficient places, children who score within the 'normal' band, but who may be displaying subtle behavioural changes or experiencing difficulties, as gleaned in the multi-disciplinary meeting, are also suitable for a place. In this way, Pyramid Club is both a targeted intervention for children at the SEND level of SEMH, and a preventative intervention for CYP at risk.

Psychological Background

Pyramid Club aims to drive behaviour change through a competence enhancement framework, rooted in Positive Psychology (Jayman et al., 2019a). Positive Psychology can be described as a strengths-based view of what contributes to people's "flourishing" or "optimal functioning" (Gable & Haidt, 2005). The intervention focuses on CYP's strengths to support them to construct solutions, rather than to focus on child deficits. Jayman et al. (2019a, para. 5) detail "The Pyramid ethos rests on four tenets of healthy child development (Kellmer-Pringle, 1980); these reflect the Pyramid Club experience for children: praise and recognition, love and security, new experiences,

and responsibility". Club leaders are trained to deliver therapeutic activities (see Table 1) and in behavioural techniques to positively reinforce club attendees and act as positive role models. Jayman et al. (2019a) highlight that CYP engage in group-based therapeutic activities (see Table 1) which support confidence and resilience building and improve social skills. Jayman et al. (2019b) investigated the behaviour change procedures and drivers which underpin Pyramid Club and identified key contextual procedures of: setting, delivery and content criteria, and key change drivers of: demonstration and practice, social reward, social support and goal setting.

Review Question

Pyramid Club is a low-cost intervention (individual schools can purchase the intervention for around £1000, including licencing, training for up to 16 club leaders and resources and materials, with this cost reducing for multiple schools under a council or academy chain) and over 33,000 children have already attended clubs across the UK (University of West London, n.d.). Given this, the author feels a synthesis of the literature to determine effectiveness is warranted.

Therefore, the current review aims to determine: how effective is Pyramid Club at reducing internalising behaviours and increasing prosocial skills for children and young people aged 7-14 who have identified socio-emotional needs?

Critical Review of the Evidence Base

Literature Search

A systematic literature search was conducted in January 2022. Google Scholar was used for an initial scoping search and to identify key search terms. Following this, a

systematic literature search was conducted using four electronic databases: Google Scholar, Web of Science for its broad focus on multidisciplinary literature, ERIC (EBSCO) for its focus on literature in the field of Education and PsycINFO for its focus on literature in the field of Psychology. The search terms used to identify studies in relation to the review question are outlined in Table 2.

Table 2
Search Terms

| Intervention | | Participant | | Outcome |
|------------------------|-----|--------------|-----|-----------------------|
| | | child OR | | |
| “Pyramid Club” OR | | children OR | | “social emotional” OR |
| “Pyramid Project” OR | AND | “school age” | AND | “socio emotional” OR |
| “Pyramid Intervention” | | OR | | “emotional wellbeing” |
| | | students | | |

Note: quotation marks are used to search for an exact phrase and group words together, “OR” is used to find one term or the other. Search terms were then combined using “AND” to find papers with all of the search terms.

The initial search generated a total of 122 results. From these, 10 duplicates were removed and 112 screened by title and abstract. 94 were excluded according to various inclusion and exclusion criteria (see Table 3), leaving 18 for full text screening. The full text screen resulted in 13 studies being removed (see Appendix 2 for details), leaving a total of 5 eligible for review (see Table 4). Figure 1 illustrates a flow diagram of the search strategy.

Figure 1

PRISMA Flow Diagram of Search Strategy

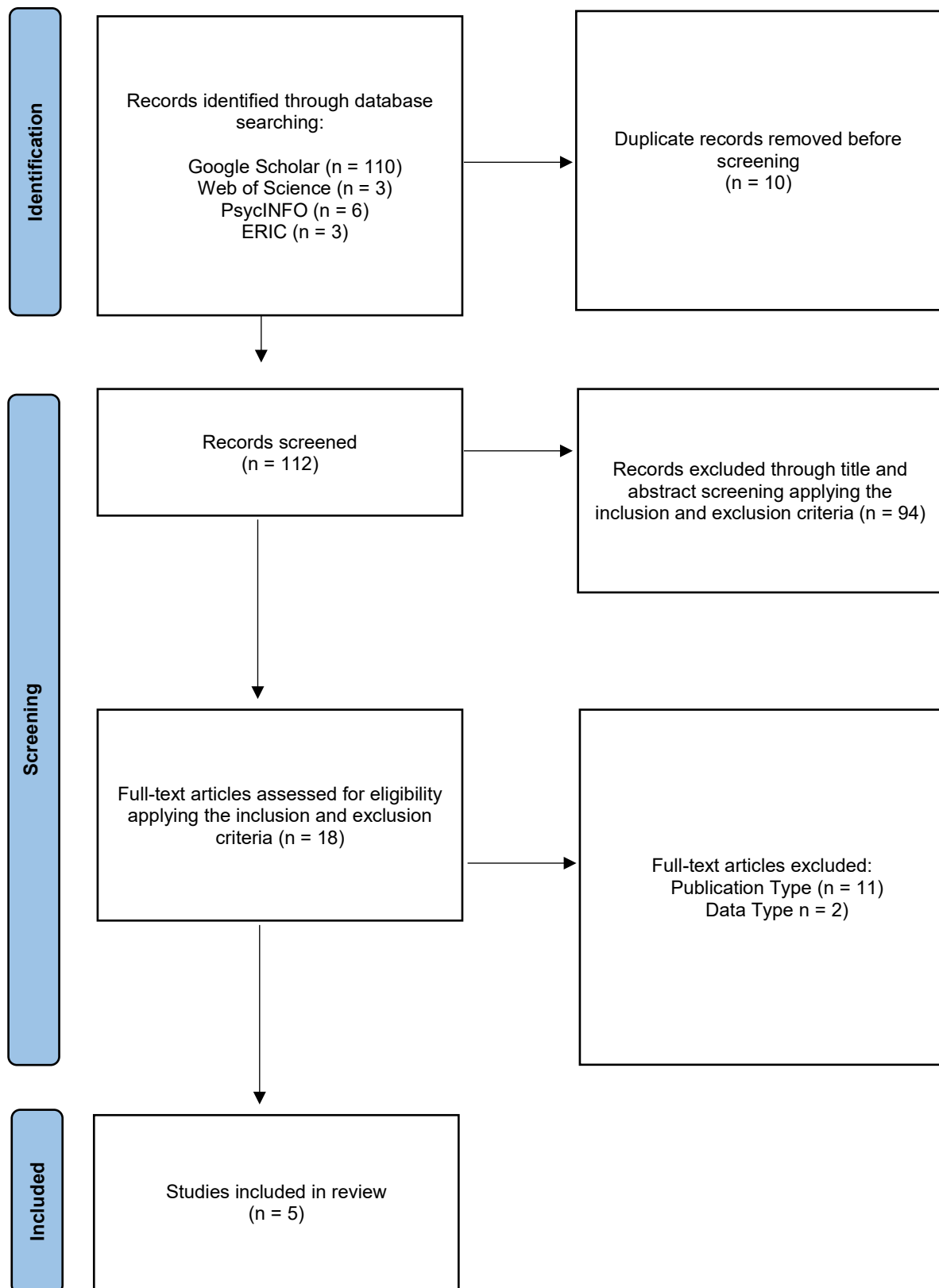


Table 3

Inclusion and Exclusion Criteria for Current Review

| Number | Criterion | Inclusion Criteria | Exclusion Criteria | Rationale |
|--------|----------------------|--|--|---|
| 1. | Publication Type | The articles must be accessible in a peer reviewed journal | The articles had not been published in an accessible peer review journal | To ensure studies are of a high calibre |
| 2. | Data Type | The study used purely quantitative methods to evaluate the impact of Pyramid Club or quantitative data within a mixed methods design was included. | The study used purely qualitative methods to evaluate the impact of Pyramid Club or qualitative data within a mixed methods design was excluded. | Petticrew and Roberts (2003) define qualitative data as low down in their typology of evidence for effectiveness evaluation |
| 3. | Study Design | Intervention studies that use an experimental or quasi-experimental research design to evaluate the impact of Pyramid Club | Articles that do not use an experimental or quasi-experimental method or review findings from previously conducted studies | The review aims to evaluate the effectiveness of Pyramid Club and therefore relies on original empirical data |
| 4. | Participants | The study involved children between 7-14 years old with identified social-emotional needs | Children younger than 7 or older than 14 years and with no identified social-emotional needs | The Pyramid Club intervention targets children aged 7-14 with identified social-emotional needs |
| 5. | Intervention | The study used the Pyramid Club intervention | The study does not include the Pyramid Club intervention | This review seeks to explore the effectiveness of this intervention |
| 6. | Setting | Conducted in a school-based setting | Conducted outside of a school-based setting | This review seeks to explore the effectiveness of school-based interventions only |
| 7. | Outcome and Measures | The study must evaluate pre and post measures of | The study does not evaluate pre and post | The review question is interested in any |

| | | | | |
|----|----------|--------------------------------------|--|--|
| | | internalising behaviours. | measures of internalising behaviours | change in identified internalising behaviours |
| 8. | Language | Publications are produced in English | Publications are not produced in English | The author's primary language is English and translation services were not available |

Table 4

Studies Included in the Current Review

| Included Studies | |
|------------------|--|
| 1 | McKenna, Á. 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. <i>Child and Adolescent Mental Health, 19</i> (1), 9-15. |
| 2 | Cassidy, T., McLaughlin, M., & Giles, M. (2014). Group-based intervention to improve socioemotional health in vulnerable children. <i>Journal of Psychology & Clinical Psychiatry, 1</i> (7), 00045. |
| 3 | Jayman, M., Ohl, M., Hughes, B., & Fox, P. (2019). Improving socio-emotional health for pupils in early secondary education with Pyramid: A school-based, early intervention model. <i>British Journal of Educational Psychology, 89</i> (1), 111-130. |
| 4 | Ohl, M., Fox, P., & Mitchell, K. (2013). Strengthening socio-emotional competencies in a school setting: Data from the Pyramid project. <i>British Journal of Educational Psychology, 83</i> (3), 452-466. |
| 5 | 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. <i>Child and Adolescent Mental Health, 13</i> (3), 115-121. |

Mapping the Field

The studies included in the review shared commonalities in terms of the research designs used, methods of evaluation and intervention implementation to investigate the effectiveness of Pyramid Clubs on socio-emotional well-being in children aged 7-14 years old. Details of mapping the field can be found in Appendix 3.

Weight of Evidence

Gough’s Weight of Evidence Framework (WoE; Gough, 2007) defines three dimensions for weighting and critically appraising research: methodological quality (WoE A; see Appendices 4, 5 and 6), methodological relevance (WoE B; see Appendix 7) and relevance of the topic to the review question (WoE C; see Appendix 8). Finally, these three weightings are combined and averaged to give an overall weighting to a study (WoE D). Table 5 summarises WoE ratings.

Table 5

WoE Summary

| Study | WoE A | WoE B | WoE C | WoE D |
|-----------------------|-----------------|-----------------|-----------------|-----------------|
| McKenna et al. (2014) | 0 (very low) | 2.2 (medium) | 1.6 (medium) | 1.3 (low) |
| Cassidy et al. (2014) | 1 (low) | 2.6 (high) | 1.8 (medium) | 1.8 (medium) |
| Jayman et al.(2019a) | 3 (high) | 2.4 (medium) | 2.4 (medium) | 2.6 (high) |
| Ohl et al.(2013) | 0 (very low) | 1.8 (medium) | 2.6 (high) | 1.5 (medium) |
| Ohl et al. (2008) | 1 (low) | 1.6 (medium) | 2 (medium) | 1.5 (medium) |

Note. WoE D ratings are defined as <1.5 as ‘low’, ≥ 1.5 and < 2.5 as ‘medium’ and ≥ 2.5 as ‘high’.

Participants

There was a total of 971 participants across the 5 studies, with a sample size range of 82 – 375. The age of participants ranged from 7-14, although only one study was conducted with secondary age pupils (Jayman et al, 2019a), whilst the other four were conducted in primary schools. However, as the intervention Pyramid Club is targeted towards 7-14-year olds, this difference was not reflected in the WoE C criteria.

In addition, three studies reported the gender split for participants with complete data for analysis: a total of 40 males and 48 females (24 males and 33 females in intervention condition; McKenna et al., 2014); 122 males and 172 females (62 males and 100 females in intervention condition; Cassidy et al., 2014); finally, 52 males and 74 females (26 males and 40 females in intervention condition; Jayman et al., 2019a). One study reported gender split before participant exclusions for data analysis: 200 males and 185 females (46 males and 57 females in intervention condition; Ohl et al., 2013) and one study only reported the total sample gender split (52 males and 51 females) and size of the intervention condition ($n = 42$), but not the gender split within the intervention condition and after participant exclusions (Ohl et al., 2008).

In the three studies that did report gender split in intervention conditions, there were always more females than males. This may reflect findings that female gender has been found to be a risk factor for internalising difficulties (Mendelson et al., 2008) and clinical disorders associated with internalising symptoms are more prevalent in females (Lewinsohn et al., 1993).

Two studies were conducted in schools in Northern Ireland (McKenna et al., 2014; Cassidy et al., 2014), one in schools in England and Wales (Jayman et al., 2019a) and two in schools in England (Ohl et al., 2013; Ohl, et al., 2008). This was not reflected in the WoE C criteria as all the studies were conducted in the UK, where the educational systems and stressors are similar. The number of schools in each study ranged from 4 – 13, with an average of 7.8 schools. Similarly, this was not reflected in WoE C as it was deemed all of the studies had a sufficient minimum number of schools for generalisability. UK based evidence is preferable for EPs working in a UK context.

Conversely, participant characteristics were deemed important for generalisability and were therefore included in WoE C criteria. Participant characteristics included: age, school year, gender, ethnicity, socio-economic status (SES; as determined by FSM eligibility) and location of school (i.e. urban or rural). Two studies defined 3 or less participant characteristics (McKenna et al., 2014; Cassidy et al., 2014) and subsequently received a 'low' WoE C rating in this criterion. One study defined 4 participant characteristics (Ohl et al., 2008) and received a WoE C 'medium' rating in this criterion. Finally, two studies defined 5 or more participant characteristics and therefore received a WoE C 'high' rating (Jayman et al., 2019a; Ohl et al., 2013).

Study Design

All 5 studies in the present review were quasi-experimental. Petticrew and Roberts (2003) rate quasi-experimental as second best to answering effectiveness questions. As such, each study received a WoE B criterion rating of 'medium'. Similarly, each study employed a control group which received no intervention or attention placebo, earning a WoE B criterion rating of 'medium'. However, whilst three studies had non-

matched control groups which, earning them a WoE B of 'medium' (McKenna et al., 2014; Ohl et al., 2013; Ohl, et al., 2008), two studies employed control groups which were matched either on participant characteristics (Jayman et al., 2019a) or through a waiting list control condition (Cassidy et al., 2014), earning a WoE B rating of 'high'.

Finally, Eliopoulos et al. (2005) detail justification of research design and recognition of potential limitations as criterion for effective quasi-experimental designs. One study made no justification for the use of research design, earning a WoE C criterion rating of 'low' (Ohl et al., 2008), two made justification but did not detail the limitations of this design, earning a WoE C rating of 'medium' (Jayman et al., 2019a; Ohl et al., 2013) and two made both justification and reference to potential limitations, earning a WoE C rating of 'high' (McKenna et al., 2014; Cassidy et al., 2014).

Intervention Implementation

All studies in the current review screened for and implemented the intervention according to the Pyramid Club manual (Pyramid, 2007; as cited in Ohl et al., 2013). As a result, this aspect was not reflected in WoE C criteria. However, studies differed in their reporting of the club leader (reflected in WoE A) and intervention fidelity and dosage threshold (reflected in WoE C). Three studies reported the club leader as: trained co-ordinators (Cassidy et al., 2014), school support staff and community volunteers (Jayman et al., 2019a) and trained volunteers (Ohl et al., 2008). Two studies did not report the club leader (McKenna et al., 2014; Ohl et al., 2013).

Dosage threshold is set as a minimum of 70% attendance in the Pyramid Club manual. Studies which monitored intervention fidelity and excluded data for analysis where

dosage threshold was not met, achieved 'high' WoE C ratings (Jayman et al., 2019a; Ohl et al., 2013). Studies which monitored intervention fidelity but not dosage threshold received a WoE C rating of 'medium' (Cassidy et al., 2014). Finally, studies which did not monitor intervention fidelity or dosage threshold resulted in a WoE C rating of 'low' (McKenna et al., 2014).

Outcome Measures

Gersten et al. (2005) details how multiple outcome measures should ideally be used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance. In addition, follow up measures can provide information about long-term intervention effects. Two studies tested only at pre and post intervention with no follow up, which was reflected in a 'low' WoE B criterion rating (Ohl et al., 2013; Ohl, et al., 2008), one study tested at pre, post and follow up but used only one measure (teacher report SDQ; McKenna et al., 2014), resulting in 'medium' ratings. Lastly, two studies not only tested at pre, post and follow up time points, but used more than one measure, utilising the teacher and self-report SDQ (Cassidy et al., 2014; Jayman et al., 2019a), which resulted in a favourable WoE C rating.

In addition, as the present review is interested in any change in internalising behaviours, WoE C ratings favoured studies that reported and analysed the SDQ sub domains ('internalising' and 'externalising'), rather than just the Total Difficulties (TD) scores. Reporting only TD scores makes isolating the effect of Pyramid Club on internalising behaviours alone more difficult, for which one study was penalised (Ohl et al., 2008). One study reported and analysed the 'internalising' sub scales only

(McKenna et al., 2014). Two studies reported and analysed both 'internalising' and 'externalising' domains, resulting in a more robust analysis of the impact of Pyramid Club on difficulties (Cassidy et al., 2014; Jayman et al., 2019a), which in turn received 'high' WoE C ratings.

Moreover, the present review is interested in internalising behaviours at the SEN level. Therefore, WoE C criteria reflected greater emphasis on using the parent report SDQ and at least the teacher or self-report SDQ. This is because Goodman et al. (2000) found that the parent report SDQ offers more predictive value for emotional disorders than the teacher or self-report SDQ and more than one informant is the most robust application of the SDQ. However, no studies in the present review utilised the parent report SDQ which was reflected in WoE C ratings. Two studies used at least the teacher and self-report SDQ scales, receiving 'medium' ratings for doing so (Cassidy et al., 2014; Jayman et al., 2019a). Three studies were penalised for using only one informant SDQ scale (teacher rated), receiving 'low' ratings (McKenna et al., 2014; Ohi et al., 2013; Ohi et al., 2008).

Finally, studies that reported descriptives for participants' SDQ banding categories and banding shifts following intervention were given higher WoE C ratings. SDQ scores have been found to be associated with clinical diagnoses (Goodman et al., 2000). The current review is interested in CYP with SEN, therefore participants with initial SDQ bands pertaining to SEN, that is 'abnormal' (clinical diagnosis) and 'borderline' (at risk), are more relevant. Two studies did not report the number of participants in each SDQ banding and banding shifts (Cassidy et al., 2014; Jayman et al., 2019a). Whereas, three studies reported the number of participants who fell into SDQ banding categories

and any observed shifts in these bandings following intervention, resulting in 'high' ratings (McKenna et al., 2014; Ohl et al., 2013; Ohl et al., 2008).

Findings and Effect Sizes

Table 6 summarises the effect sizes for each study, evaluated using Cohen's (1988) threshold. Effect sizes were calculated where possible or converted to Cohen's *d* using Psychometrica (Lenhard & Lenhard, 2016). Effect sizes reported for each study in Table 6 are related to interaction effects.

All studies found that Pyramid Club had a significant impact on the reduction in SDQ rated domains, with four studies isolating this effect to internalising behaviours (ranging from small-large effect sizes) and three of these studies demonstrating sustained reduction (McKenna et al., 2014; Cassidy et al., 2014; Jayman et al., 2019a). One study found a significant impact of Pyramid Club with a large effect size, but did not further analyse TD scores to isolate the effect to particular subscales (Ohl et al., 2008).

Of the three studies which also measured and analysed 'externalising' subscales, two found no significant change in 'conduct problems' and 'hyperactivity' (Jayman et al., 2019a; Ohl et al., 2013), whilst one found a sustained medium effect on teacher rated 'hyperactivity' and a sustained small effect on self-report 'conduct' (Cassidy et al., 2014). These findings are not surprising, given that Pyramid Club targets children with 'internalising' difficulties and screens out those with 'externalising' difficulties.

McKenna et al. (2014) found medium effect sizes for the impact of Pyramid Club on teacher rated SDQ 'internalising' scores. The study was also one of three that reported SDQ banding and banding shifts sustained over time following the intervention (see Table 6), with 19 participants scoring 'high' (a combination of 'abnormal' and 'borderline' banding) at pre-test, reducing to 4 at post-test and increasing slightly, at 12-week follow up to 6 participants. However, the research received the lowest overall WoE D rating and received a 'very low' score for its methodological quality. Therefore, results should be interpreted with caution.

Cassidy et al. (2014) found large effect sizes for the impact of Pyramid Club on teacher rated and self-rated internalising behaviours and prosocial skills (ranging from $d = 0.84 - 1.22$). However, the study's results write up was deemed to not be very clear, which contributed a 'low' WoE A rating. The study did not report SDQ or SDQ banding shifts which was reflected in WoE C. Overall, the study received the second highest WoE D of 'medium'.

Ohl et al. (2013) found a large effect size for the reduction of teacher rated 'emotional symptoms' and a medium effect size for reduction in 'peer problems', sustained at follow up. SDQ banding shifts revealed 23 participants measuring at 'abnormal' and 10 at 'borderline' pre-intervention, reduced to 11 participants at 'abnormal' and 12 at 'borderline' post-test. The study achieved a medium WoE D rating, mainly hindered by its 'very low' score in methodological quality.

Ohl et al. (2008) found a large effect size for the impact of Pyramid Club on SDQ TD scores. However, they did not perform further analysis on the subscales to isolate this effect further. They did look at SDQ banding and banding shifts and found 15 participants measuring at the 'abnormal' level and 12 participants at the 'borderline' level pre-test, was reduced to 3 at 'abnormal' and 10 at 'borderline' following the intervention.

Jayman et al. (2019a) received the highest overall WoE D rating of 'high' for their research and was the only study conducted in a secondary school setting. Whilst the researchers did not report individuals' SDQ banding or banding shifts, their research was of a high methodological quality, methodological relevance and topic relevance for the review. They found 'large' effect sizes for the impact of Pyramid Club on teacher rated internalising behaviours, which was sustained at follow up, and small effect sizes for self-report internalising behaviours, which was not explored at follow up. Small effect sizes may be explained by social desirability bias present within self-report measures.

Table 6

Summary of Effect Sizes

| Study | Sample Size | Measure | Sub Scales | Pre-Post | | Follow Up | | WoE D |
|-----------------------|-------------|-------------------|--------------------|-----------------|---------------------|-----------------|---------------------|--------------|
| | | | | Effect Size (d) | Label (Cohen, 1988) | Effect Size (d) | Label (Cohen, 1988) | |
| McKenna et al. (2014) | 82 | Teacher rated SDQ | Emotional Symptoms | | | .74*** | Medium | 1.3 (low) |
| | | | Peer Problems | | | .63*** | Medium | |
| | | | Prosocial Skills | | | ns | | |
| Ohl et al. (2013) | 375 | Teacher rated SDQ | Emotional Symptoms | .87*** | Large | | | 1.5 (medium) |
| | | | Peer Problems | .62** | Medium | | | |
| | | | Prosocial Skills | .50* | Medium | | | |
| | | | Conduct Problems | ns | | | | |
| | | | Hyperactivity | ns | | | | |
| Ohl et al. (2008) | 94 | Teacher rated SDQ | Total Difficulties | 1.34*** | Large | | | 1.5 (medium) |

| Study | Sample Size | Measure | Sub Scales | Pre-Post | | Follow Up | | WoE D | |
|-----------------------|-------------|-------------------|--------------------|--------------------------|---------------------|--------------------------|---------------------|--------------|--|
| | | | | Effect Size (<i>d</i>) | Label (Cohen, 1988) | Effect Size (<i>d</i>) | Label (Cohen, 1988) | | |
| Cassidy et al. (2014) | | Teacher rated SDQ | Emotional Symptoms | | | 1.12*** | Large | 1.8 (medium) | |
| | | | Peer Problems | | | 0.84*** | Large | | |
| | | | Prosocial Skills | | | 1.03*** | Large | | |
| | | | Conduct Problems | | | ns | | | |
| | | | Hyperactivity | | | 0.51*** | Medium | | |
| | | | Total Difficulties | | | 0.67*** | Medium | | |
| | | | | | | | | | |
| | | Self-rated SDQ | Emotional Symptoms | | | | 1.22*** | Large | |
| | | | Peer Problems | | | | 0.91*** | Large | |
| | | | Prosocial Skills | | | | 1.12*** | Large | |
| | | | Conduct Problems | | | | 0.35** | Small | |
| | | | Hyperactivity | | | | ns | | |
| | | | Total Difficulties | | | | 0.62*** | Medium | |
| | | | | | | | | | |

| Study | Sample Size | Measure | Sub Scales | Pre-Post | | Follow Up | | WoE D |
|-----------------------|-------------|-------------------|--------------------|--------------------------|---------------------|--------------------------|---------------------|------------|
| | | | | Effect Size (<i>d</i>) | Label (Cohen, 1988) | Effect Size (<i>d</i>) | Label (Cohen, 1988) | |
| Jayman et al. (2019a) | 126 | Teacher rated SDQ | Emotional Symptoms | .88*** | Large | 0.80*** | Large | 2.6 (high) |
| | | | Peer Problems | .99*** | Large | 0.90*** | Large | |
| | | | Prosocial Skills | .44* | Small | ns | | |
| | | | Conduct Problems | ns | | | | |
| | | | Hyperactivity | ns | | | | |
| | | | Total Difficulties | .98*** | Large | 1.05*** | Large | |
| | | Self-rated SDQ | Emotional Symptoms | 0.38* | Small | | | |
| | | | Peer Problems | 0.44* | Small | | | |
| | | | Prosocial Skills | ns | | | | |
| | | | Conduct Problems | ns | | | | |
| | | | Hyperactivity | ns | | | | |
| | | | Total Difficulties | 0.41** | Small | | | |

Note. *** $p < .001$, ** $p < .01$, * $p < 0.5$, ns = not significant.

Subscales ‘Emotional Symptoms’ and ‘Peer Problems’ total to make ‘Internalising’ scores. Subscales ‘Conduct Problems’ and ‘Hyperactivity’ total to make ‘Externalising’ scores. Total Difficulties scores are the total combination of ‘Emotional Symptoms’, ‘Peer Problems’, ‘Conduct Problems’ and ‘Hyperactivity’ subscales. ‘Prosocial Skills’ is a strengths-based subscale. Pyramid Club aims to reduce internalising scores and increase prosocial skills scores.

WoE D ratings are defined as < 1.5 as ‘low’, ≥ 1.5 and < 2.5 as ‘medium’ and ≥ 2.5 as ‘high’.

Table 7

Summary of Participants’ SDQ Banding Shifts

| Study | Measure | Sub Scales | SDQ Banding (pre) | SDQ Banding (post) | SDQ Banding (follow up) |
|-----------------------|-------------------|--------------------|----------------------------|--------------------|-------------------------|
| McKenna et al. (2014) | Teacher rated SDQ | Emotional Symptoms | 19 P’s ^a ‘High’ | 4 P’s ‘High’ | 6 P’s ‘High’ |
| | | | 38 P’s ‘Normal’ | 53 P’s ‘Normal’ | 51 P’s ‘Normal’ |
| | | Peer Problems | 13 P’s ‘High’ | 2 P’s ‘High’ | 3 P’s ‘High’ |
| | | | 44 P’s ‘Normal’ | 55 P’s ‘Normal’ | 56 P’s ‘Normal’ |
| Cassidy et al. (2014) | N/A | N/A | N/A | N/A | N/A |
| Jayman et al. (2019a) | N/A | N/A | N/A | N/A | N/A |
| Ohl et al. (2013) | Teacher rated SDQ | Total Difficulties | 23 P’s ‘Abnormal’ | 11 P’s ‘Abnormal’ | |

| Study | Measure | Sub Scales | SDQ Banding (pre) | SDQ Banding (post) | SDQ Banding (follow up) |
|-------------------|-------------------|--------------------|---------------------|---------------------|-------------------------|
| | | | 10 P's 'Borderline' | 12 P's 'Borderline' | |
| | | | 69 P's 'Normal' | 79 P's 'Normal' | |
| Ohl et al. (2008) | Teacher rated SDQ | Total Difficulties | 15 P's 'Abnormal' | 3 P's 'Abnormal' | |
| | | | 12 P's 'Borderline' | 10 P's 'Borderline' | |
| | | | 15 P's 'Normal' | 29 P's 'Normal' | |

Note. 'Abnormal' (clinical threshold) and 'Borderline' (at risk) SDQ banding have been grouped together as 'High' where the study did not distinguish between them, 'normal' reflects the SDQ banding of 'normal'. ^a P's = participants

Conclusions and Recommendations

Conclusions

The aim of the present review was to explore the effectiveness of Pyramid Club in supporting the socio-emotional well-being of children between 7 and 14 years old. From the 5 studies reviewed, one had a WoE D rating of 'low' (mainly pertaining to its 'very low' methodological quality; McKenna et al., 2014), with the other four ranging from medium to high, therefore results from this study in particular should be interpreted with caution.

In conclusion, evidence for the effectiveness of Pyramid Club on reducing internalising behaviours ('emotional symptoms' and 'peer problems') is strong. Across 4 studies that isolated their analyses on the internalising subscales of 'total difficulties', the average effect size was $d = 0.85$ (large) for 'emotional symptoms' and $d = 0.72$ (medium-high) for 'peer problems'. However, evidence for the effectiveness of Pyramid Club on increasing 'prosocial skills' is more mixed, with one study finding no significant improvement (McKenna et al., 2014), one study finding a small ($d = .44$) effect on teacher rated behaviours, but no significant self-rated increase (Jayman et al., 2019a), one study finding a medium effect size ($d = .50$; Ohl et al., 2013) and finally, one finding large ($d = >1$) effect sizes for both teacher and self-rated behaviours (Cassidy et al., 2014). In the one study that looked at overall 'total difficulties' scores, they found a large effect size ($d = 1.34$) for the impact of Pyramid Club (Ohl et al., 2008), although this effect cannot be isolated to particular aspects of socio-emotional well-being as it can in the other 4 studies.

Furthermore, in the two studies that examined the impact of Pyramid Club on category shifts in the severity of 'total difficulties', they found Pyramid attendees showed greater levels of improvement in post intervention banding than the control group (Ohl et al., 2013; Ohl et al., 2008). Finally, one study found a 68% decrease in 'emotional symptoms' from 'high' (i.e. 'abnormal' and 'borderline') category banding to 'normal' and a 77% decrease in 'peer problems' from 'high' category banding to 'normal' in Pyramid Club attendees (McKenna et al., 2014).

Limitations and Future Research

The studies in the present review were limited in that none of them utilised the parent version of the SDQ for effectiveness evaluation. Research has found the parent version to have the highest predictive value for emotional disorders in CYP and multiple informant measures is the most robust application of the SDQ (Goodman et al., 2000). Additionally, systems around the child should work in partnership to promote CYP's well-being; when schools and parents collaborate, children's learning, well-being and outcomes in school are improved (Oostdam & Hooge, 2013). Therefore, future research in this area should utilise at least the parent version, alongside other measures, to triangulate CYP's difficulties.

Furthermore, all of the studies utilised no intervention control groups. Future research with more stringent methodological processes, such as matched control groups that received an alternative intervention or an attention placebo, would help to isolate the impact of Pyramid Club over and above existing interventions. Finally, research into Pyramid Club thus far has been carried out by a particular cluster of researchers. To

reduce potential bias, future research using independent research teams would be preferable.

Implications for EP Practice

The present review highlights the role schools can play in promoting socio-emotional well-being in CYP. Pyramid Club appears to be a low cost, yet effective intervention for CYP who present with internalising difficulties. Additionally, there is the added benefit of being able to upskill existing school staff to deliver the intervention 'in house'. Upskilling existing members of school staff will give them the tools to utilise elements of Pyramid Club practice to other contexts. Pyramid Club has been shown to improve internalising behaviours for children across primary and secondary school settings, across a number of schools, highlighting its generalisability. These findings put EPs in a unique position to recommend an intervention that has its evidence base in a UK context.

References

- Bryant, A., Guy, J., Holmes, J., Astle, D., Baker, K., Gathercole, S., ... & CALM Team. (2020). The strengths and difficulties questionnaire predicts concurrent mental health difficulties in a transdiagnostic sample of struggling learners. *Frontiers in Psychology*, 3125.
- Campbell, D. T., & Stanley, J. C. (2015). *Experimental and Quasi-experimental Designs for Research*. Ravenio books.
- Cassidy, T., McLaughlin, M., & Giles, M. (2014). Group-based intervention to improve socioemotional health in vulnerable children. *Journal of Psychology & Clinical Psychiatry*, 1(7), 00045.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum
- Department for Education & Department of Health. (2015). *Special educational needs and disability code of practice: 0 to 25 years: Statutory guidance for organisations which work with and support children and young people who have special educational needs or disabilities*.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/398815/SEND_Code_of_Practice_January_2015.pdf
- Eliopoulos, G. M., Harris, A. D., Lautenbach, E., & Perencevich, E. (2005). A systematic review of quasi-experimental study designs in the fields of infection control and antibiotic resistance. *Clinical Infectious Diseases*, 41(1), 77-82.
- Gable, S. L., & Haidt, J. (2005). What (and why) is positive psychology? *Review of General Psychology*, 9(2), 103-110.

- 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.
- Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*(5), 791-799.
- Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(4), 400-403.
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry, 177*(6), 534-539.
- Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry, 7*(3), 125-130.
- Goodwin, R. D., Fergusson, D. M., & Horwood, L. J. (2004). Early anxious/withdrawn behaviours predict later internalising disorders. *Journal of Child Psychology and Psychiatry, 45*(4), 874-883.
- 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.
- Hiscock, H., Bayer, J. K., Lycett, K., Ukoumunne, O. C., Shaw, D., Gold, L., ... & Wake, M. (2012). Preventing mental health problems in children: the Families in Mind population-based cluster randomised controlled trial. *BMC Public Health, 12*(1), 1-9.

- Jayman, M., Ohl, M., Hughes, B., & Fox, P. (2019a). Improving socio-emotional health for pupils in early secondary education with Pyramid: A school-based, early intervention model. *British Journal of Educational Psychology*, 89(1), 111-130.
- Jayman, M., Ohl, M., & Fox, P. (2019b). Improving wellbeing for pupils in early secondary education with Pyramid Club: a qualitative study investigating behaviour change drivers. *The Psychology of Education Review*, 43(2).
- Kellmer-Pringle, M. (1980). *The needs of children: A personal perspective*. London, UK: Hutchinson
- Lenhard, W. & Lenhard, A. (2016). Computation of effect sizes. Retrieved from: https://www.psychometrica.de/effect_size.html. Psychometrica. DOI: 10.13140/RG.2.2.17823.92329
- Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III—R disorders in high school students. *Journal of Abnormal Psychology*, 102(1), 133.
- McKenna, Á. 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.
- Mendelson, T., Kubzansky, L. D., Datta, G. D., & Buka, S. L. (2008). Relation of female gender and low socioeconomic status to internalizing symptoms among adolescents: A case of double jeopardy?. *Social Science & Medicine*, 66(6), 1284-1296.

- 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*(3), 452-466.
- 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.
- Oostdam, R., & Hooge, E. (2013). Making the difference with active parenting; forming educational partnerships between parents and schools. *European Journal of Psychology of Education, 28*(2), 337-351.
- Petrides, K. V. (2009). Psychometric properties of the Trait Emotional Intelligence Questionnaire. In C. Stough, D. H. Saklofske, and J. D. Parker, *Advances in the Assessment of Emotional Intelligence*. New York: Springer. DOI: 10.1007/978-0-387-88370-0_5
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology & Community Health, 57*(7), 527-529.
- Thomson, K. C., Richardson, C. G., Gadermann, A. M., Emerson, S. D., Shoveller, J., & Guhn, M. (2019). Association of childhood social-emotional functioning profiles at school entry with early-onset mental health conditions. *JAMA Network Open, 2*(1), e186694-e186694.
- University of West London. (n.d.) *Pyramid Club Evidence*. Retrieved February 10, 2022, from <https://www.uwl.ac.uk/business/pyramid-clubs-schools/pyramid-club-evidence>

Appendices

Appendix 1 – SDQ Information

The SDQ (Goodman, 1997) was designed to be a brief emotional and behavioural screening questionnaire for CYP aged 3-17 years old. There are several informant versions, with teacher and self-report versions being utilised in the studies in the present review. There also exists variations of a parent version. The SDQ is broken down into 5 sub scales, each containing 5 items on psychological attributes: 'emotional symptoms', 'conduct problems', 'hyperactivity/inattention', 'peer problems' and 'prosocial skills'. The latter sub scale is scored as a 'strengths' item, whilst the other four are seen as 'difficulties' items. Scores on sub scales can be combined to create a 'total difficulties' scores ('emotional symptoms', 'conduct problems', 'hyperactivity/inattention' and 'peer problems') or further categorised to create 'internalising problems' ('emotional symptoms' and 'peer problems') and 'externalising problems' ('conduct problems' and 'hyperactivity/inattention'). Scores can be divided into bandings, derived from normative UK data (Meltzer et al., 2000), which include: 'abnormal' (10% of the population), 'borderline' (10% of the population) and 'normal' (80% of the population).

Appendix 2 – Literature Exclusion

Table 8

Studies Excluded from Analysis After Screening

| Reference | Reason for Exclusion | Inclusion/ Exclusion Criteria Number |
|---|--|--------------------------------------|
| Jayman, M., Ohl, M., & Fox, P. (2019). Improving wellbeing for pupils in early secondary education with Pyramid Club: a qualitative study investigating behaviour change drivers. <i>The Psychology of Education Review</i> , 43(2). | The study included only qualitative data | 2 |
| Lyons, R., & Woods, K. (2012). Effective Transition to Secondary School for Shy Children: A case study using ‘Pyramid’ group work. <i>Educational and Child Psychology</i> , 29(3), 8-26. | The study was published from a thesis but the text was not accessible. | 1 |
| Jayman, M., Fox, P., Ohl, M., & Hughes, B. (2017). Beyond evidence-based interventions: implementing an integrated approach to promoting pupil mental wellbeing in schools with pyramid club. <i>Education and Health</i> , 35(4), 70-74. | Not a peer reviewed article (report) | 1 |
| Jayman, M., Fox, P., Ohl, M., & Hughes, B. (2017). Improving socio-emotional health and school performance for pupils in early secondary education with Pyramid: a school-based, early intervention model. | Not a peer reviewed article (conference paper) | 1 |
| Jayman, M., Ohl, M., Hughes, B., & Fox, P. (2014). The impact of a group based intervention on the social and emotional competencies of young people in early secondary education: a pilot study. | Not a peer reviewed article (research poster) | 1 |
| Jayman, M., Ohl, M., Hughes, B., & Fox, P. (2015). The Pyramid Project at the University of West London. | Not a peer reviewed article (presentation) | 1 |
| Jayman, M. (2018). The impact Pyramid Club has on socio-emotional health and school performance for pupils in early secondary education: a mixed methods evaluation study. | Not a peer reviewed article (conference paper) | 1 |

| | | |
|---|---|----------|
| <p>Ohl, M. (2021). Pyramid Club: building skills for healthy friendships and relationships in a digital age. <i>Supporting New Digital Natives: Children's Mental Health and Wellbeing in a Hi-Tech Age</i>, 23.</p> | <p>Not a peer reviewed article (book chapter)</p> | <p>1</p> |
| <p>Jayman, M., Ohl, M., Hughes, B., & Fox, P. (2015). Evaluating the impact of the Pyramid intervention on the emotional health and school performance of students in early secondary education.</p> | <p>Not a peer reviewed article (conference paper)</p> | <p>1</p> |
| <p>Jayman, M., Ohl, M., Hughes, B., & Fox, P. (2015). Improving the socio-emotional health of young people in early secondary education: preliminary findings from a study of the Pyramid intervention project. <i>Resilience and Health in a Fast-Changing World</i>, 109-120.</p> | <p>Not a peer reviewed article (book chapter)</p> | <p>1</p> |
| <p>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).</p> | <p>The study included qualitative evaluation in a retrospective follow up</p> | <p>2</p> |
| <p>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>. The University of Manchester (United Kingdom).</p> | <p>Not a peer reviewed article (thesis)</p> | <p>1</p> |
| <p>Cassidy, T., McLaughlin, M., & Giles, M. (2015). Socio-emotional health in school children: An emotion-focused intervention. <i>International Journal of School and Cognitive Psychology</i>, 2(4), 1-7.</p> | <p>Could not access text</p> | <p>1</p> |

Appendix 3 – Mapping the Field

Table 9

Mapping the Field

| Author, Date and Title | Study Design | Country and Setting | Participants | Intervention Details | Outcome Measures | Main Findings |
|---|---|--|--|---|---|--|
| <p>Author: McKenna et al. (2014)</p> <p>Title: 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</p> | <p>Design: quasi-experimental</p> <p>Groups: non-equivalent group design</p> <p>Intervention: 31 children</p> <p>Control: 57 children</p> <p>Data Collection: Pre, post and 12-week follow up</p> | <p>Country: Northern Ireland</p> <p>Setting: 7 primary schools</p> | <p>Sample Size: 82</p> <p>Age: 7-8 years (school year 4 – England year 3 equivalent)</p> <p>Gender: 40 males, 48 females</p> | <p>Intervention: Pyramid Club</p> <p>Duration: 10 weeks</p> <p>Total Lessons: 10 sessions</p> <p>Lesson Duration: 90 minutes</p> <p>Frequency: x1 per week</p> <p>Club Leader: not reported</p> <p>Control Condition: class children who were screened for the intervention but did not meet threshold (no intervention).</p> | <p>Socio-emotional Status: Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) provided teacher ratings on 3/5 scales:</p> <p>Emotional Symptoms and Peer Problems ('difficulties' items) and Pro-social Behaviour ('strengths' item)</p> | <p>Findings: Pyramid Club participants showed greater teacher reported reductions than the comparison group at follow up on the Emotional Symptom and Peer Problem factors. There was no significant effect on Prosocial Skills.</p> <p>Teacher rated SDQ banding for Emotional Symptoms shifted from 19 to 4 to 6 participants' scoring 'high'^a at follow up. Peer Problems banding shifted from 13 to 2 to 3 participants' scoring 'high' at follow up.</p> |

| Author, Date and Title | Study Design | Country and Setting | Participants | Intervention Details | Outcome Measures | Main Findings |
|--|---|---|--|---|---|---|
| <p>Author: Cassidy et al. (2014)</p> <p>Title: Group-based intervention to improve socio-emotional health in vulnerable children</p> | <p>Design: quasi-experimental</p> <p>Groups: non-equivalent group design</p> <p>Intervention: 162 children</p> <p>Control: 132 children</p> <p>Data Collection: pre, post and 12 weeks follow ups</p> | <p>Country: Northern Ireland</p> <p>Setting: 13 primary schools</p> | <p>Sample Size: 294</p> <p>Age: 11 years (school year 7 – England year 6 equivalent)</p> <p>Gender: 122 males, 172 females</p> | <p>Intervention: Pyramid Club</p> <p>Duration: 10 weeks</p> <p>Total Lessons: 10 lessons</p> <p>Lesson Duration: 90 minutes</p> <p>Frequency: x1 per week</p> <p>Club leader: trained co-ordinator</p> <p>Control Condition: waiting list control group (no intervention)</p> | <p>Socio-emotional Status: Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) provided teacher and self-report ratings on 5/5 sub scales: Emotional Symptoms, Conduct, Hyperactivity, Peer Problems ('difficulties' items) and Pro-social Behaviour ('strengths' items).</p> | <p>Findings: Pyramid Club participants showed greater reductions at follow up in teacher and self-report ratings than the comparison group for Emotional Symptoms and Peer Problems and increased teacher and self-report Prosocial Skills ratings.</p> |
| <p>Author: Jayman et al.(2019a)</p> <p>Title: Improving socio-emotional health for pupils</p> | <p>Design: quasi Experimental</p> | <p>Country: England and Wales</p> | <p>Sample Size: 126</p> <p>Age: 11-14 years (school</p> | <p>Intervention: Pyramid Club</p> <p>Duration: 10 weeks</p> | <p>Socio-emotional Status: Strengths and Difficulties Teachers Questionnaire</p> | <p>Findings: Pyramid Club participants showed greater reductions in teacher and self-report ratings than the comparison group for</p> |

| Author, Date and Title | Study Design | Country and Setting | Participants | Intervention Details | Outcome Measures | Main Findings |
|---|--|--|---|--|--|--|
| in early secondary education with Pyramid: A school-based, early intervention model | Groups: matched groups Intervention: 66 Control: 60 Data Collection: pre-test, post-test and 12 month follow up | Setting: 8 secondary schools | year 7, 8 and 9) Gender: 52 males, 74 females | Total Lessons: 10 Lesson Duration: 90 minutes Frequency: x1 per week Club leader: school support staff and community volunteers Control Condition: matched (age, gender & FSM) group (no intervention) | (SDQ; Goodman, 1997) and the self-report version (Goodman et al., 1998) on 5/5 subscales: Emotional Symptoms, Conduct, Hyperactivity, Peer Problems ('difficulties' items) and Pro-social Behaviour ('strengths' items). | Emotional Symptoms and Peer Problems which was maintained at teacher report follow up. Prosocial Skills teacher, not self-report, ratings increased, but was not maintained at teacher report follow up. |
| Author: Ohi et al. (2013) Title: Strengthening socio-emotional competencies in a school setting: | Design: quasi-experimental Groups: non-equivalent groups Intervention: 102 | Country: England Setting: 7 Primary schools | Sample Size: 375 Age: 7-8 years (year 3) Gender: 200 males, 185 | Intervention: Pyramid Club Duration: 10 weeks Total Lessons: 10 lessons | Socio-emotional status: The Strengths and Difficulties Teachers Questionnaire (SDQ; Goodman, 1997) was used on 5/5 | Findings: Pyramid Club participants showed greater reductions in teacher report ratings than the comparison group for 'emotional symptoms', 'peer problems' and increased 'prosocial skills'. |

| Author, Date and Title | Study Design | Country and Setting | Participants | Intervention Details | Outcome Measures | Main Findings |
|---|--|--|--|---|---|---|
| Data from the Pyramid project | Control: 273 Data: pre-test and post-test (12-week post intervention) | | females (only reported before 10 participant data exclusions – final gender split unknown) | Lesson Duration: 90 minutes Frequency: x1 per week Club leader: not reported Control Condition: class children who were screened for the intervention but did not meet threshold (no intervention) | subscales: Emotional Symptoms, Conduct, Hyperactivity, Peer Problems ('difficulties' items) and Pro-social Behaviour ('strengths' items). | Pre-post teacher rated SDQ banding shifted from 23 to 11 participants' scoring 'abnormal', 10 to 12 participants' scoring 'borderline' and 69 to 79 participants' scoring 'normal' post intervention. |
| Author: Ohl et al.(2008) Title: The Pyramid Club Primary School-Based Intervention: Evaluating the | Design: quasi-experimental Groups: Intervention: 42 Control: 52 | Country: England Setting: 4 primary schools | Sample Size: 94 Age: 8-9 years Gender: 54 males, 51 females (only | Intervention: Pyramid Club Duration: 10 weeks Total Lessons: 10 Lesson Duration: 90 minutes | Socio-emotional status: The Strengths and Difficulties Teachers Questionnaire (SDQ; Goodman, 1997) was used on 4/5 | Findings: Pyramid Club participants showed greater reductions in their teacher reported SDQ TD scores than the comparison group. Pre-post teacher rated SDQ banding shifted |

| Author, Date and Title | Study Design | Country and Setting | Participants | Intervention Details | Outcome Measures | Main Findings |
|--|------------------------------|---------------------|--|--|---|---|
| Impact on Children's Social-Emotional Health | Data: pre-test and post-test | | reported before 11 participant data exclusions – final gender split unknown) | Frequency: x1 per week Club Leader: trained volunteers Control Condition: class children who were screened for the intervention but did not meet threshold (no intervention) | subscales: Emotional Symptoms, Conduct, Hyperactivity, Peer Problems ('difficulties' items) to give a Total Difficulties Score. | from 15 to 3 participants' scoring 'abnormal', 12 to 10 participants' scoring 'borderline' and 15 to 29 participants' scoring 'normal'. |

Note. ^a 'Abnormal' (clinical threshold) and 'Borderline' (at risk) SDQ banding have been grouped together as 'High' where the study did not distinguish between them, 'normal' reflects the SDQ banding of 'normal'.

Appendix 4 – Amendments to Coding Protocol

An amended version of the Gersten et al. (2005) coding protocol for group experimental and quasi-experimental research in special education was used. The following question amendments were made (where a line through indicates the removal of text):

- 1) Essential quality indicator question 3 – Was sufficient information given characterizing the interventionists or teachers provided? ~~Did it indicate whether they were comparable across conditions?~~

This was justified as Pyramid Club is a school-based intervention; studies in this review used either non-intervention or wait list control groups that did not receive an alternative intervention and therefore information characterising interventionists in these control conditions was not relevant.

- 2) Desirable quality indicator question 2 – 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?~~

This was because it was not deemed relevant in relation to the nature of the Pyramid Club intervention. As Pyramid Club is a school-based intervention, (which can be implemented both during school hours or after school), and all studies employed at least the teacher rated SDQ measure, it is difficult in this case to achieve scorers blind to study conditions.

Appendix 5 – Example of a Completed Coding Protocol

McKenna, Á. 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.

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?

Yes

No

Unknown

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

Yes

No

Unknown

3. Was sufficient information given characterizing the interventionists or teachers provided?

Yes

No

Unknown

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

4. Was the intervention clearly described and specified?

Yes

No

Unknown

5. Was the fidelity of implementation described and assessed?

Yes

No

Unknown

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

Yes

No

Unknown

Quality Indicators for Outcome Measures

7. Were multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance?

- Yes
- No
- Unknown

8. Were outcomes for capturing the intervention's effect measured at the appropriate times?

- Yes
- No
- Unknown

Quality Indicators for Data Analysis

9. 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

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

- Yes
- No
- Unknown

Desirable Quality Indicators

1. 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%?

- Yes
- No
- Unknown

2. 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

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

- Yes

- No
- Unknown

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

- Yes
- No
- Unknown

5. 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

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

- Yes
- No
- Unknown

7. Did the research report include actual audio or videotape excerpts that capture the nature of the intervention?

- Yes
- No
- Unknown

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

- Yes
- No
- Unknown

Total Essential = 5

Total Desirable = 1

WoE A Rating = 0 (very low)

Appendix 6 – WoE A (Methodological Quality)

Table 10 demonstrates the WoE A threshold criteria used in the current review as suggested by Gersten et al. (2005). As Gersten and researchers only identified threshold criteria for ‘very high’ and ‘acceptable’ research, thresholds were introduced to measure ‘low’ and ‘very low’ rated studies. Table 11 highlights the WoE A ratings for the 5 studies in the present review.

Table 10

WoE A Rating Thresholds based on the Gersten et al. (2005) Coding Protocol

| WoE A Rating | Criteria |
|-----------------|---|
| 3 (high) | Study meets at least 9 essential criteria and at least 4 desirable criteria |
| 2 (medium) | Study meets at least 9 essential criteria and at least 1 but less than 4 desirable criteria |
| 1 (low) | Study meets at least 6 but fewer than 9 essential criteria |
| 0 (very low) | Study meets less than 6 essential criteria |

Table 11

Scores from the WoE A Coding

| Criteria Category | McKenna et al. (2014) | Cassidy et al. (2014) | Jayman et al.(2019a) | Ohl et al. (2013) | Ohl et al. (2008) |
|----------------------------------|-----------------------|-----------------------|----------------------|-------------------|-------------------|
| Participation description (/3) | 1 | 2 | 3 | 1 | 2 |
| Intervention Implementation (/3) | 1 | 2 | 2 | 2 | 2 |
| Outcome Measures (/2) | 1 | 2 | 2 | 0 | 1 |
| Data Analysis (/2) | 1 | 2 | 2 | 2 | 2 |
| Total Essential (/10) | 5 | 8 | 9 | 5 | 7 |
| Total Desirable (/8) | 1 | 3 | 4 | 3 | 2 |
| WoE A Rating | 0 (very low) | 1 (low) | 3 (high) | 0 (very low) | 1 (low) |

Appendix 7 – WoE B (Methodological Relevance)

Table 12

WoE B Rating Thresholds

| Criterion | Weighting | Rationale |
|--------------------------------------|---|--|
| A. Study Design | 3 Randomised Control Trials | Petticrew and Roberts (2003) outline in their typology of evidence the ranking of study designs to address questions of effectiveness |
| | 2 Quasi-experimental and Cohort studies | |
| | 1 Qualitative research, Survey, Case-control studies or Non-experimental evaluations | |
| B. Control Group Conditions | 3 Use of a control group which receives an alternative intervention or an attention placebo control | Control groups should be as stringent as possible in order to isolate the effect of the independent variable on the dependent variable (Campbell & Stanley, 2015) |
| | 2 Use of a control group which receives no intervention | |
| | 1 No use of a control group | |
| C. Control Group Matching | 3 Control groups were matched | Control groups should be as stringent as possible in order to isolate the effect of the independent variable on the dependent variable (Campbell & Stanley, 2015) |
| | 2 Control groups were not matched | |
| | 1 No use of a control group | |
| D. Outcome Measures | 3 Use of more than one pre-test, post-test and follow up measures | Multiple outcome measures should ideally be used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalised performance. In addition, including follow up measures can provide information of long-term intervention effects (Gersten et al., 2005). |
| | 2 Use of just one pre-test and post-test and follow up measures | |
| | 1 Use of pre-test(s) and post-test(s) but no follow up measure(s) | |

| | | | |
|---|---|---|--|
| E. Justification of Design and Recognition of Limitations | 3 | Justification of research design is made with reference to the potential limitations of that design | Eliopoulos et al.2005) detail justification of research design and recognition of potential limitations as criterion for effective quasi-experimental designs. |
| | 2 | Justification for the use of research design is made | |
| | 1 | No justification for the use of research design is made | |

Table 13

Scores from the WoE B Coding

| Study | Criteria A Rating | Criteria B Rating | Criteria C Rating | Criteria D Rating | Criteria E Rating | WoE B Rating |
|-----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------|
| McKenna et al. (2014) | 2 | 2 | 2 | 2 | 3 | 2.2 (medium) |
| Cassidy et al. (2014) | 2 | 2 | 3 | 3 | 3 | 2.6 (high) |
| Jayman et al. (2019a) | 2 | 2 | 3 | 3 | 2 | 2.4 (medium) |
| Ohl et al. (2013) | 2 | 2 | 2 | 1 | 2 | 1.8 (medium) |
| Ohl et al.(2008) | 2 | 2 | 2 | 1 | 1 | 1.6 (medium) |

Note. WoE B ratings ≥ 2.5 are defined as 'high', ≥ 1.5 and < 2.5 as 'medium' and < 1.5 as 'low'.

Appendix 8 – WoE C (Topic Relevance)

Table 14

WoE C Criteria Thresholds

| Criterion | Weighting | Rationale |
|-----------------------------|---|---|
| A. Evidence Gathering | 3 The study uses the parent report version of the SDQ <i>and</i> at least the teacher <i>or</i> self-report SDQ | Goodman et al. (2000) found that the parent report SDQ offers more predictive value for emotional disorders than teacher report and self-report data (which are about equally as useful as each other). The use of more than one informant is the most robust application of the SDQ. |
| | 2 The study uses at least the teacher <i>and</i> self-report SDQ | |
| | 1 The study uses the teacher <i>or</i> self-report SDQ | |
| B. Evidence Reporting | 3 The study reported and analysed at least internalising <i>and</i> externalising SDQ sub scale data. | The present review is interesting in looking at the effectiveness of Pyramid Club, an intervention which targets internalising behaviours. Reporting both internalising and externalising SDQ subscale data allows for comparisons to monitor whether there were any unexpected positive or adverse significant changes in CYP's externalising scores as a result of the intervention. Where SDQ sub scale scores were not reported and analysed, it is harder to isolate the effect of Pyramid Club on internalising behaviours alone. |
| | 2 The study reported and analysed internalising SDQ sub scale data only | |
| | 1 The study reported and analysed Total Difficulties score data of the SDQ only. | |
| C. Descriptive Data | 3 The study reported the number of participants in SDQ banding categories <i>and</i> banding shifts following the Pyramid Club intervention | SDQ scores have been found to be positively associated with clinical diagnoses (Goodman et al., 2000). As the current review is interested in CYP with SEN, studies that reported descriptives on the number |
| | 2 The study reported the number of participants in | |

| | | |
|--|--|--|
| | SDQ banding categories but <i>not</i> banding shifts following the Pyramid Club intervention | of participants in SDQ banding and banding shifts of scores pertaining to SEN (i.e. 'abnormal' - clinical diagnosis and 'borderline' - at risk) are more relevant to the aims of the review. |
| | 1 The study did not report the number of participants in SDQ banding categories or shifts in banding | |
| D. Participant Characteristics ^a | 3 ≥ 5 participant characteristics are included | The more detailed a study reports their participant characteristics, the more generalisable the findings are. |
| | 2 4 participant characteristics are included | |
| | 1 ≥ 3 participant characteristics are included | |
| E. Intervention Fidelity and Dosage Threshold | 3 The study excluded data from analysis when they attended less than 7/10 Pyramid Club sessions <i>and</i> intervention fidelity was monitored | Monitoring intervention fidelity ensures that the intervention is being delivered to the standard it is supposed to be delivered. Dosage threshold is recommended in the Pyramid Club manual as it reflects the requisite time for children to form a group which they may benefit from (Ohl, 2009). This is set at 70%. |
| | 2 The study monitored intervention fidelity but did <i>not</i> monitor dosage threshold | |
| | 1 The study did not monitor dosage threshold <i>and/or</i> fidelity | |

Note. ^a Participant characteristics: age, school year, gender, ethnicity, socio-economic status (determined by FSM eligibility) and location of school (i.e. urban or rural).

Table 15

WoE C Ratings

| Study | Criteria A Rating | Criteria B Rating | Criteria C Rating | Criteria D Rating | Criteria E Rating | WoE C Rating |
|-----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|--------------|
| McKenna et al. (2014) | 1 | 2 | 3 | 1 | 1 | 1.6 (medium) |
| Cassidy et al. (2014) | 2 | 3 | 1 | 1 | 2 | 1.8 (medium) |
| Jayman et al. (2019a) | 2 | 3 | 1 | 3 | 3 | 2.4 (medium) |

| | | | | | | |
|-------------------|---|---|---|---|---|---------------|
| Ohl et al. (2013) | 1 | 3 | 3 | 3 | 3 | 2.6 (high) |
| Ohl et al.(2008) | 1 | 1 | 3 | 2 | 3 | 2 (medium) |

Note. WoE C ratings ≥ 2.5 are defined as 'high', ≥ 1.5 and < 2.5 as 'medium' and < 1.5 as 'low'.