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

[illegible]

[illegible]

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

Figure 1

PRISMA Flow Diagram of Search Strategy

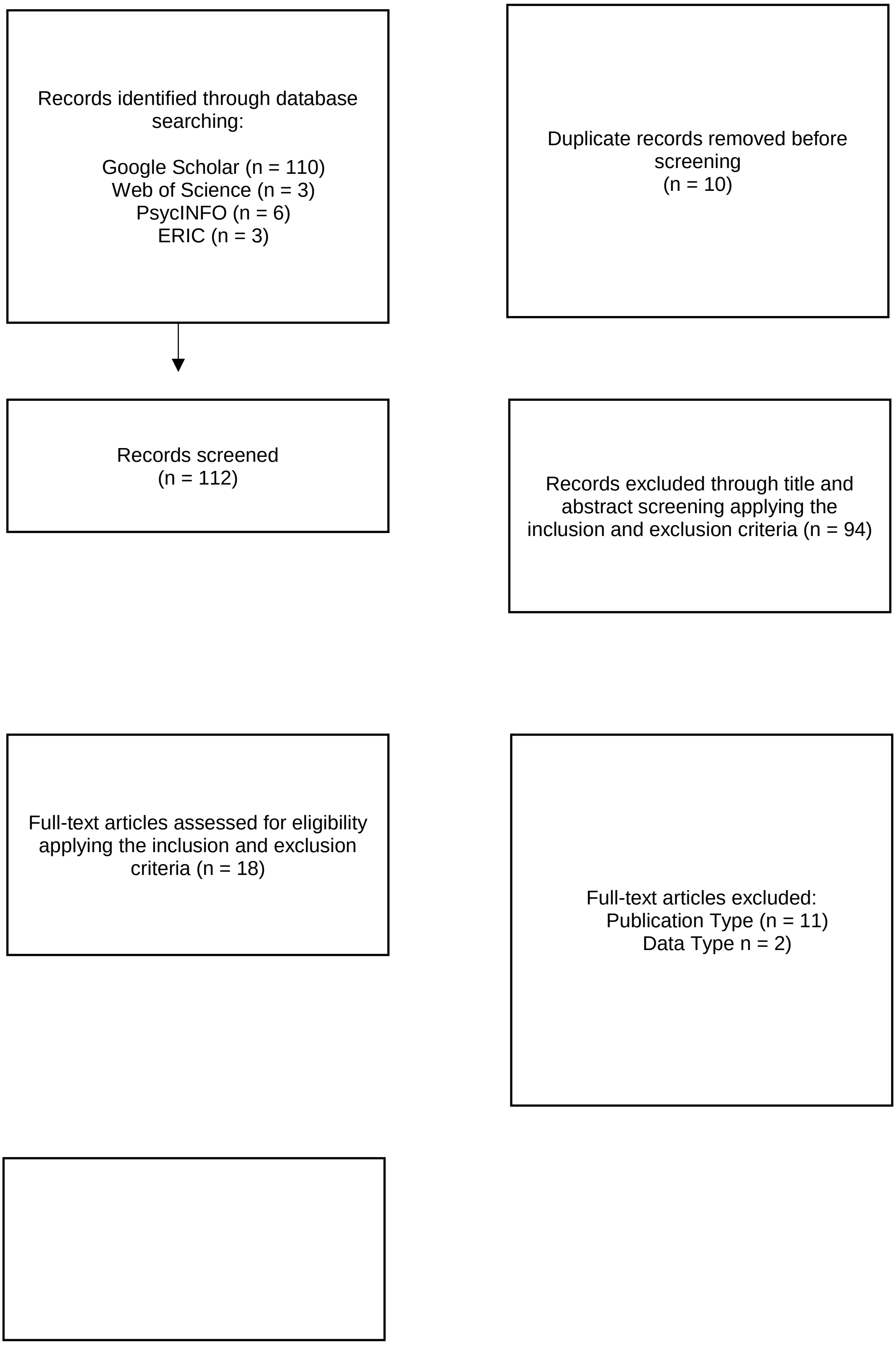

Table 3

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			

		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. *Child and Adolescent Mental Health*, *19*(1), 9-15.
2	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.
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. *British Journal of Educational Psychology*, *89*(1), 111-130.
4	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.
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. *Child and Adolescent Mental Health*, *13*(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

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. Td[oc)4 (i)6 (o)]TJethni,

ab,lk ConvcdeSE

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’

(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; Ohl et al., 2013; Ohl 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

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)

				Pre-Post		Follow Up		
Study	Sample Size	Measure	Sub Scales	Effect Size (*d*)	Label (Cohen, 1988)	Effect Size (*d*)	Label (Cohen, 1988)	WoE D
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		

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)

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