

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

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

How effective are school-based Attribution Retraining interventions in reducing peer-directed aggression in school children?

Summary

The purpose of this review is to evaluate the effectiveness of school-based Attribution Retraining (AR) interventions in reducing peer-directed aggression in school-aged children. Attributions are inferences that people make about others' feelings and beliefs, in order to understand the intent behind their behaviour. Dodge and Coie (1987) argue that maladaptive attributions can lead to situations of causal uncertainty being perceived as hostile, which can lead to aggressive behaviour. AR has been suggested as a means of identifying these attributions and making them more adaptive, to reduce aggressive behaviour (Vassilopoulos et al., 2015). This review sought to evaluate the efficacy of AR interventions in reducing hostile attributions and, in turn, aggressive behaviour within a school context. A comprehensive literature search returned five studies, which met the review's inclusion criteria. These were evaluated using Gough's (2007) Weight of Evidence (WoE) Framework. A meta-analysis was carried out and showed a large combined effect size when comparing AR interventions to the 'no-intervention' control group in each study. This, along with the medium to large overall quality of the reviewed studies, suggests school-based AR interventions can be recommended as evidence-based practice (Gersten et al., 2005) in reducing aggression in school-age populations. Limitations of the review, along with

recommendations for educational psychology practice and for future research, are discussed.

Introduction

Attribution Theory

Attribution Theory (AT) proposes that individuals seek to determine the cause of observable behaviours and outcomes (Heider, 1958), and attribute feelings, beliefs, and intentions to others in an attempt to explain what they have observed. An individual may believe the cause to be: internal or external; controllable or uncontrollable; stable or unstable (Weiner, 1985). Their interpretation will impact their thoughts and, in turn, their subsequent behaviour (Frederickson & Cline, 2007).

In relation to aggression, AT has been used to aid understanding of how an individual's perception of an event can lead to aggressive behaviour. Crick and Dodge (1994) believe that a child's behaviour results from six sequential steps of processing: (1) encoding of social cues; (2) interpreting social cues; (3) selection of goals; (4) generation of response; (5) response evaluation; and (6) enacting of behaviour, with item (2) being considered fundamental. This is because, if social cues are misinterpreted as ambiguous peer provocations, it can affect the child's wellbeing (Dodge, 2006) and potentially lead to aggressive behaviour.

Research demonstrates that some aggressive individuals may have a hostile attribution bias (Nasby et al., 1980), which leads to situations of causal ambiguity being perceived as provocative (Van Bockstaele et al., 2020). The causes of this have been attributed to executive functioning deficits (Verhoef et al., 2019); cognitive

processing difficulties (Dodge et al., 1986); traumatic life events (Frederickson & Cline, 2007); early adverse childhood experiences (Verhoef et al., 2019); harsh parenting and peer rejection (Dodge, 1980). Dodge's (2006) model of social information processing provides support for this theory. The model states that aggressive individuals are more likely to interpret others' motives as confrontational in ambiguous situations, rather than harmless or neutral.

Social Cognitive Theory states that this can lead to the development and maintenance of aggression (Verhoef et al., 2019), with individuals becoming aggressive as a way to retaliate or defend themselves. This maintenance occurs due to the aggressive individual not having the opportunity to challenge their hostile beliefs or learn more prosocial behaviour strategies. A vicious cycle can then be generated where hostile attribution bias leads to aggression and further rejection by their peer group. This can then lead the aggressive individual to perceive the world as a hostile place and this perpetuates their aggressive behaviour.

The role of hostile attribution bias in developing and perpetuating aggressive behaviour has been supported in longitudinal (Lansford et al., 2010), experimental (Lochman & Dodge, 1998), and longitudinal-experimental studies (Lochman & Wells, 2002). This provides a rationale for an intervention that educational psychologists (EPs) can recommend, which focuses on adapting attribution bias in order to reduce aggressive behaviour.

Attribution Retraining (AR)

AR aims to identify maladaptive causal attributions, challenge misconceptions and help individuals learn more adaptive ways of perceiving and interpreting themselves

and others. In this review, this means adapting a child's hostile attribution bias in order to reduce their aggressive behaviour.

AR interventions have been studied in a variety of settings, including clinics (Hilt, 2004) and schools (Robertson, 2000), and have been shown to be successful in adapting attributions relating to achievement (Chodkiewicz & Boyle, 2014), as well as behaviour (Lapointe & Legault, 2004). There are dedicated AR programs, such as the Brain Power Program (Hudley et al., 1994), as well as study-specific AR interventions (Vassilopoulos et al., 2015; Van Bockstaele et al., 2020). This means that studies on AR interventions can vary according to: environmental factors; length of the intervention; number of sessions; content and target group. They have also been carried out on whole classes (Ziegler & Heller, 2000) and smaller groups (Chodkiewicz & Boyle, 2014). Content delivery can also vary depending on the methods used, such as the use of written resources (Vassilopoulos et al., 2015) and video simulations (Van Bockstaele et al., 2020) and techniques employed can also vary and include: persuasion; motivation; problem-solving; modelling; calming exercises and the use of self-talk.

Rationale and Relevance

One of the most common forms of social difficulty among school-aged children is the exhibiting of externalising behaviours such as aggression and anger (Wilmshurst, 2009). These result in difficulties with peer relationships, poor self-concept and academic underachievement in the short term (Vassilopoulos et al., 2015) and can lead to substance misuse and criminality in the long term (Reef et al., 2011). This highlights the need for effective strategies to be put in place in order to support children in reducing these behaviours.

It is important that any intervention adopted to support children with these difficulties are evidence-based and relevant contextually. An EP can play an important role in supporting children with these difficulties through identifying areas of need, within the school and the child, and recommending appropriate interventions that match the individual context. An EP can also oversee implementation and support monitoring and evaluation.

It is important that EPs play a role in supporting these children as otherwise they can be at risk of both social and school exclusion (Reed, 1988), especially if staff do not understand that the child's behaviour could be a means of communicating that they feel threatened.

Research linking the use of AR with aggressive children dates back to the 1970s with Harris and Huang (1974) reporting that the cognitive attribution theory of emotion was important in explaining aggressive behaviour. This was supported by Graham (1997) who found that students who had been labelled 'aggressive' were biased towards perceiving peer provocation as intentional. Through the use of AR, Graham showed that those who received AR had altered attributions and emotional reactions post-training, which indicates that AR may be an approach EPs could recommend to schools. However, despite the reported success of AR in reducing aggressive behaviour, the majority of the research has been conducted outside of the UK. Therefore, there is a need to assess the true effectiveness of AR interventions before ascertaining if they should be recommended for implementation by EPs within a UK context.

Review Question

How effective are school-based Attribution Retraining interventions in reducing peer-directed aggression in school children?

Critical Review of the Evidence Base

Literature search

A systematic literature search using the Web of Science; PsycINFO (OVID); SCOPUS and ERIC (EBSCO) databases was carried out on 15th December 2020. The search terms used across all four databases are outlined in Table 1 and were confined to peer reviewed journal articles. Ancestral and citation searches were also conducted on articles selected for inclusion in the review.

Table 1: Search terms used in this systematic literature review

Search Term	Rationale
"attribution retraining" OR "attribution modification" OR "retrain attribution" OR "attribution intervention"	This review is interested in studies which used attribution retraining approaches
child OR youth OR "young people" OR adolescent OR minor OR teen OR school	This review is interested in school-based attribution retraining interventions carried out on school-aged children
aggression OR "aggressive behaviour" OR violent* OR hostile	This review is interested in studies measuring outcomes of aggressiveness and hostile attributions

Note: Truncation (*) was used to include endings of root words; speech marks (“”) ensure inclusion of exact phrases

Article Screening

Initial database searches, as well as ancestral and citation searches, yielded 216 results. After removing 27 duplicates, title and abstract screening was carried out on 189 studies to determine eligibility for inclusion in the review. At this stage the inclusion and exclusion criteria (see Table 2) were applied, and a further 174 records were excluded, leaving 15 articles for ‘full-text’ screening. Ten studies were removed at this stage (see Appendix A), leaving five studies eligible for review (see Table 3). This process is outlined in Figure 1 in the form of a flow diagram.

Table 2: Inclusion and Exclusion Criteria with Rationale

Criterion	Inclusion	Exclusion	Rationale
1. Location	Worldwide scope	No geographical location excluded	To ensure an international perspective was derived
2. Participants	All children/young people in formal education including those with SEN (age 5-18)	Adults (18+) All children under 5 and those not in formal education	To assess AR within a school context
3. Setting	AR must take place within the school the participant is attending	AR outside of the school the participant is attending	To assess AR within a school context
4. Intervention	AR is the sole intervention or is adapted in minor ways	Non-AR interventions AR not main focus	To assess impact of AR over other approaches
5. Type of Study 1	Randomised Control Trial (RCT) Studies with a ‘no intervention’ control group	Not RCT Studies with no ‘no intervention’ control group and studies with a second treatment group	RCTs are most suitable for effectiveness questions (Petticrew & Roberts, 2003)
6. Type of Study 2	Must empirically examine the	Systematic literature reviews	To review original findings

	relationship between participation in an AR intervention ^a in a school and changes in aggression ^b	Meta-analysis	
7. Language	Published in the English language	Not published in the English language	To ensure full understanding and correct interpretation
8. Type of Publication	Peer reviewed journal articles	Grey literature and non-peer reviewed literature	To ensure the methodological quality has been rigorously scrutinised
9. Outcomes	Must measure children's aggression	Outcome measures not related to aggression	To ensure construct validity of findings
10. Publication Date	The article is published on or before 15 th December 2020	The article is published after 15 th December 2020	To ensure all available relevant articles on the date of the literature search are considered for inclusion

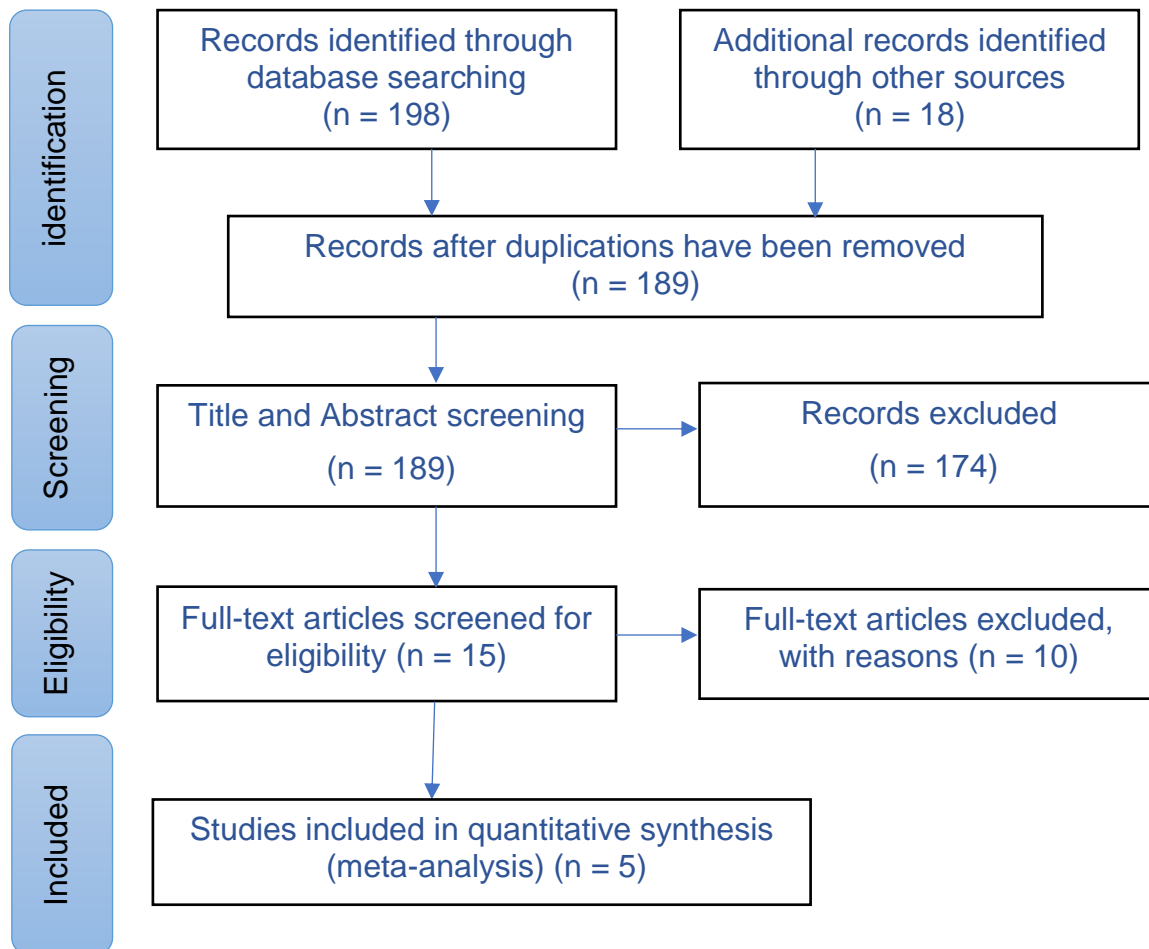
a The nature of AR is evaluated in Weight of Evidence C; if a study explicitly claimed its intervention involved AR, it met the inclusion criterion.

b The type of aggression is not specified in the inclusion criterion as there are a variety of definitions used in the literature.

Table 3: References of included studies

Studies in this review
1 Vassilopoulos, S. P., Brouzos, A. & Andreou, E. (2015). A Multi-Session Attribution Modification Program for Children with Aggressive Behaviour: Changes in Attributions, Emotional Reaction Estimates, and Self-Reported Aggression. <i>Behavioural and Cognitive Psychotherapy</i> , 43 (5), 538 – 548. https://doi.org/10.1017/S1352465814000149
2 Hudley, C., Britsch, B., Wakefield, W. D., Smith, T., Demorat, M. & Cho, S. J. (1998). An attribution retraining program to reduce aggression in elementary school students. <i>Psychology in the schools</i> , 35 (3), 271-282. <a href="https://doi.org/10.1002/(SICI)1520-6807(199807)35:3<271::AID-PITS7>3.0.CO;2-Q">https://doi.org/10.1002/(SICI)1520-6807(199807)35:3<271::AID-PITS7>3.0.CO;2-Q
3 Van Bockstaele, B., van der Molen, M. J., van Nieuwenhuijzen, M. & Salemink, E. (2020). Modification of hostile attribution bias reduces self-reported reactive aggressive behavior in adolescents. <i>Journal of Experimental Child Psychology</i> , 19 (4), 1-8. https://doi.org/10.1016/j.jecp.2020.104811
4 Hudley, C. & Graham, S. (1993). An Attributional Intervention to Reduce Peer-Directed Aggression among African-American Boys. <i>Child Development</i> , 64 (1), 124-38. https://doi.org/10.1111/j.1467-8624.1993.tb02899.x
5 Abdulmalik, J., Ani, C., Ajuwon, A. J. & Omigbodun, O. (2016). Effects of problem-solving interventions on aggressive behaviours among primary school pupils in Ibadan, Nigeria. <i>Child and Adolescent Psychiatry and Mental Health</i> , 10 (31), 1-10. https://doi.org/10.1186/s13034-016-0116-5

Figure 1: Flow diagram to depict the Literature Search and Article Screening process based on PRISMA Statement Recommendations (Moher et al., 2010)



Mapping the Field

In order to be included, the reviewed studies all needed to be randomised control trials (RCTs). This is because Petticrew and Roberts (2003) stated that this was the most reliable research design for assessing the effectiveness of an intervention such as AR. Despite having similar research designs, all studies varied in relation to their participant inclusion criteria, outcome measures, intervention duration and delivery.

Weight of Evidence (WoE)

In order to effectively evaluate the quality and relevance of the studies identified, Gough's (2007) Weight of Evidence (WoE) framework was used, which considers: the methodological quality (WoE A), methodological relevance (WoE B) and topic relevance (WoE C) of each study.

An adapted version of Gersten et al.'s (2005) coding protocol was used to assess methodological quality (WoE A) as it was designed for use with studies that employ group experimental designs.

The author used Petticrew and Roberts (2003) 'typology of evidence' to develop a protocol which analysed the strengths and weaknesses of each study's methodological relevance (WOE B).

A protocol designed by the author was used for WoE C, to assess each study's topic relevance in relation to the review question. It specifically considered: the intervention focus; the outcome measures; the intervention setting; intervention implementation and the instructor used for each study.

WoE A, B and C were given equal weighting and the sum of these were averaged and rounded to the nearest whole number to give an overall Weight of Evidence D (WoE D). This indicates the strength of evidence in each study in relation to the research question posed. A summary of WoE ratings for each study is outlined in Table 4.

The criteria and rationale for all WoE ratings are outlined in Appendix B. The adapted Gersten et al.'s (2005) coding protocol is outlined in Appendix C, with rationale for changes, and completed examples of the WoE A, B and C coding protocols are given in Appendix D.

Table 4: Summary of WoE for each study

Authors	WoE A: Methodological quality	WoE B: Methodological relevance	WoE C: Topic relevance	WoE D: Overall weight of evidence
Vassilopoulos et al. (2015)	3 High	3 High	2 Medium	3 High
Hudley et al. (1998)	3 High	3 High	3 High	3 High
Van Bockstaele et al. (2020)	3 High	3 High	2 Medium	3 High
Hudley & Graham (1993)	3 High	3 High	3 High	3 High
Abdulmalik et al. (2016)	1 Low	3 High	2 Medium	2 Medium

Note: WoE ratings are described as: 'High' = 3; 'Medium' = 2; and 'Low' = 1. WoE A, B and C were given equal weighting and the sum of these were averaged and rounded to the nearest whole number to give an overall Weight of Evidence D (WoE D).

Participants

243 participants were included across the five studies. Each study varied in sample size, ranging from 34 participants (Vassilopoulos et al., 2015) to 89 participants (Hudley et al., 1998). The age of participants also varied from 8-16, with two studies focusing on primary age students (Hudley et al., 1998; Hudley & Graham, 1993); two focusing on middle-school aged children (Vassilopoulos et al., 2015; Abdulmalik et al., 2016) and one focusing on secondary aged children (Van Bockstaele et al., 2020). Participants were male only in all but one study (Vassilopoulos et al., 2015).

Participants were identified as aggressive from a variety of peer sociometric nominations and teacher rating scales across all five studies with three studies also including non-aggressive participants (Hudley & Graham, 1993; Hudley et al., 1998; Vassilopoulos et al., 2015) in order to reduce any perceived stigma from participants being involved in the intervention. In these studies, the non-aggressive participants' data was removed from subsequent analyses. All included studies reported that there were no significant differences between the intervention and control groups in relation to age, gender, or emotion reaction estimates. Participants across the five studies had a variety of ethnic backgrounds and were based in different geographical locations. These include: Greece (White European) (Vassilopoulos et al., 2015); America (African-American and Latino) (Hudley et al., 1998); Holland (White European) (Van Bockstaele et al., 2020); America (African-American) (Hudley & Graham, 1993); and Nigeria (Black African) (Abdulmalik et al., 2016).

Setting

In all five studies, the intervention was delivered in schools, as per the inclusion criteria. One study took place within a classroom setting (Abdulmalik et al., 2016) with the remainder carried out elsewhere in the school. This was taken into consideration in the WoE C for setting with Abdulmalik et al.'s (2016) study receiving a higher weighting compared to the other studies. This is because carrying out the intervention within the classroom better enables participants to generalise the skills they have learnt in the intervention to the wider school environment. However, because a specific subset of students had been chosen across a number of different year groups, they were unlikely to be educated within the same class. Therefore, the remaining four studies were given 'medium' weightings as withdrawing students for targeted interventions like AR is typical in a school context.

Research Design

Randomised controlled trials (RCTs) are considered the most appropriate research design for 'effectiveness' questions (Petticrew & Roberts, 2003) and so were in this review's inclusion criteria. The design used in each study was evaluated using the WoE A and B ratings. All of the studies were RCTs, so received a high WoE B.

All of the studies had an AR experimental group and four of the studies had a test-retest control group (Hudley et al., 1998; Hudley & Graham, 1993; Vassilopoulos et al., 2015; Van Bockstaele et al., 2020). The fifth study had a wait-list control group (Abdulmalik et al., 2016). The existence of a 'no-intervention' condition decreased the likelihood that the positive effects observed were due to factors external to the experimental condition (Barker et al., 2015). However, it may have been useful to have an 'active' control condition across all five studies as well so that the comparative benefits of the AR intervention could be assessed relative to other aggression-reducing interventions.

Once the sample was derived, the participants were randomly allocated to either the experimental or control group. The use of random allocation reduces the likelihood of selection bias and allows for a direct comparison to occur between conditions. It also ensures that the data demonstrates a more valid representation of how the wider population would respond to AR (Barker et al., 2015), which justifies the high WoE B rating this type of design, and the studies in this review, received.

Intervention Analysis

Each study varied in terms of the number of sessions; the length of the intervention; the group size; the intervention program and the outcome measures used, as can be seen in Table 5. The content of each intervention is summarised in Table 6.

The intervention content, implementation and outcome measures will be explored in more detail over the next two sections.

Intervention Content and Implementation

As can be seen in Tables 5 and 6 below, each study differed in terms of intervention content and implementation.

Three out of the five studies (Vassilopoulos et al., 2015; Hudley et al., 1998; Van Bockstaele et al., 2020) provided only the title of the instructor with no other details about their background, training or experience, which lowered their scores within both the WoE A and C ratings. The remaining two provided detailed information on the instructor and so were awarded a higher score, which improved their overall WoE A rating. However, Abdulmalik et al. (2016) did not use a teacher to implement the AR intervention therefore still received a rating of 2 for WoE C.

In addition to this, two out of the five studies (Hudley et al., 1993; Abdulmalik et al., 2016) did not provide sufficient information about the intervention content and implementation in order to enable replicability, which lowered their scores within their WoE A rating. The remaining three studies either used a recognised program, which they referenced, or provided sufficient information for replicability and so gained a higher score within the WoE A rating for intervention implementation.

In terms of WoE C, all studies, except Abdulmalik et al. (2016), used solely AR interventions and therefore received a WoE C rating of 3 for the category 'interventions'. Abdulmalik et al. (2016) received a 2 as in this study it was combined with problem-solving and calming techniques, which made it harder to identify whether it was the AR or the other aspects which led to a reduction in aggression.

The intervention took place in a school in all of the included studies, however, only Abdulmalik et al. (2016) intervention took place in a classroom therefore all of the studies received a WoE C of 2 for 'setting' except Abdulmalik et al. (2016), which received a 3. This is because it was thought generalisability would more likely to be achieved within a classroom environment.

Finally, WoE C also looked at 'implementation' of the AR intervention. In the Vassilopoulos et al. (2015); Van Bockstaele et al. (2020) and Abdulmalik et al. (2016) studies the intervention only involved direct attributional feedback and so these studies were given a WoE C rating of 2 for 'implementation'. As the Hudley and Graham (1993) and Hudley et al. (1998) studies used interventions, which included direct attributional feedback, instruction, practice and consolidation, these were given a WoE C of 3 for 'implementation' as it is felt these are more likely to be effective in reducing aggression.

Table 5: Table outlining the intervention content, implementation and outcome measures

Study	Number of sessions	Group size	Length of intervention	Intervention Program	Instructor(s)	Outcome measures
Vassilopoulos et al. (2015)	3	Not specified	A lesson (duration unknown) One week	Cognitive Bias Modification of Interpretations procedure (CBM-I) (created by author for the purpose of the study)	1 Research Assistant	The Aggression Scale (aggression) (Orpinas & Frankowski, 2001) An ambiguous vignette paradigm (attribution) (Vassilopoulos et al., 2009)
Hudley et al. (1998)	12	6 4 Aggressive students and 2 non-aggressive students	60 minutes Twice weekly 6 weeks	The Brain Power Program (Hudley, 1994)	Group Leaders	Social Skills Rating System (SSRS-T) (aggression) (Gresham & Elliot, 1990) Student attribution task (attribution) - designed by authors for this study
Van Bockstaele et al. (2020)	5	Individually	20 minutes 2 weeks	Computer-based AR intervention designed by authors for this study	Adult experimenters	Adapted Reactive Proactive Questionnaire (aggression) (Dutch translation) (Cima et al., 2013) Adapted Interpretation Recognition Task (attribution) (Houtkamp et al., 2017)

Hudley & Graham (1993)	12 (min. of 10 to be attended)	6 4 Aggressive students and 2 non-aggressive students	40-60 minutes Twice weekly 6 weeks	AR intervention designed by authors for this study	2 African-American female experimenters (educators with specific training) Met with curriculum developer weekly to ensure intervention fidelity	Laboratory Analogue task (aggression) - designed by authors for purpose of this study Teacher Checklist – aggression subscale (aggression) (Coie, 1990) Formal Disciplinary Referrals (aggression) – school-based logs Hypothetical peer provocation questionnaire (attribution) – designed by authors for the purpose of this study
Abdulmalik et al. (2016)	6	10	40 minutes Twice weekly for 3 weeks	Adapted Brain Power program (Hudley, 1998) contextualised for Nigerian environment, combined with problem-solving and calming techniques. Translated into Yoruba.	Clinical psychologist (fluent in Yoruba) Supervised by author to ensure intervention fidelity	Teacher assessment (aggression) qualitative observation Teacher-rated SDQ (aggression) (Goodman, 2001) Teacher Rating of Student's Aggressive Behaviour (TRAB) (aggression) –designed by authors for this study Self-rates Aggression Scale (SRAS) (aggression) - designed by authors for this study Social Cognition and Attribution Scale (SCAS) (attribution) - designed by authors for this study

Table 6: Intervention content for each study

Study	Intervention	Content
Vassilopoulos et al. (2015)	Cognitive Bias Modification of Interpretations procedure (CBM-I)	<p>Each session involved the following activity:</p> <ol style="list-style-type: none"> 1. Given 15 cards with descriptions of hypothetical social events and questions about their interpretation of peer intent in the scenario 2. Turn card over to check their response with correct response (a benign interpretation) without comment or feedback 3. Given a moment to think about the differences/similarities in their response compared to correct answer, before moving on to next card <p>Repeated for three sessions.</p>
Hudley et al. (1998)	The Brain Power Program	<p>Lesson 1: introduction to the program</p> <p>Lessons 2-6: strengthen ability to accurately detect others' intentions (variety of activities)</p> <p>Lessons 7-9: students taught to associate causally ambiguous situations as accidental</p> <p>Lessons 10-11: activity linking non-aggressive behavioural responses to ambiguously caused, negative social outcomes</p> <p>Lesson 12: review of concepts covered</p>
Van Bockstaele et al. (2020)	Computer-based AR intervention designed by authors for this study	<ol style="list-style-type: none"> 1. Presented ambiguous scenarios, on a computer screen, one sentence at a time 2. Content of the scenarios were based on the Novaco Anger Scale and Provocation Inventory (Novaco, 2003) and used by Hawkins and Cougie (2013). They were adapted for adolescents through the use of shorter sentences and simpler language 3. Last sentence had parts of the word missing e.g. inexp-rienc-d. The word always facilitated interpretation of the scenario in a positive or harmless manner. 4. Participants then answered a yes/no comprehension question, which focused on the positive outcome of the scenario. Correct answer was then presented and then next scenario presented.

10 scenarios were presented in each of the 5 sessions (50 scenarios in total). The 10 scenarios consisted of: 7 scenarios presented in writing and aurally; 2 scenarios also accompanied by pictures and one was presented as a video fragment. Adaptations were made to assist participants understanding. After the 10 scenarios were presented, a self-paced break was given.

Hudley & Graham (1993)

AR intervention designed by authors for this study

Lesson 1: Not specified

Lessons 2-6: Strengthen ability to accurately detect others' intentions (variety of activities)

Lessons 7-9: Students taught to associate causally ambiguous situations as accidental

Lessons 10-11: Activity linking non-aggressive behavioural responses to ambiguously caused, negative social outcomes

Lesson 12: Not specified

Focused on peer-directed social behaviour and used familiar playground situations appropriate for age of participants.

Abdulmalik et al. (2016)

Adapted Brain Power program (Hudley, 1998) contextualised for Nigerian environment, combined with problem-solving and calming techniques

Session 1: Introduction to programme and worked on motivational strategies to support engagement

Session 2: Taught students calming techniques

Session 3: Covered problem-solving strategies

Session 4 and 5: Attribution retraining: specifically focusing on distinguishing between wilful and accidental intent, and recognising ambiguity in interpersonal interactions

Session 6: Recap salient points

Outcome Measures

As can be seen in Table 5, each study assessed attribution and aggression, using different outcome measures, both pre- and post-assessment, with only one study carrying out a follow-up assessment at 12 months (Hudley et al., 1998).

Vassilopoulos et al. (2015) used The Aggression Scale (Orpinas & Frankowski, 2001) to measure the degree to which participants engage in overt aggression. It has good psychometric characteristics (Cronbach's alpha = .83 and .88 pre- and post-assessment) and correlated with teacher ratings $r=.33$ $p<.001$ and standardised peer nominations of aggressive behaviour $r=0.4$ $p<0.001$. The attribution measure was created by the author for the purpose of the study and had moderate to high internal consistency pre- and post- intervention. Use of standardised measures contributed to this study's high WoE A and C for outcome measures.

Teachers rated students' aggression using the teacher version of the Social Skills Rating System (Gresham & Elliot, 1990) in Hudley et al.'s (1998) study. The psychometric properties for this scale are very good. The coefficient reliability for the self-control subscale is .91, and the test-retest reliability coefficient is .80. A student attribution task was created by the authors for the purpose of this study, based on research by Graham et al. (1992), which showed that this was a reliable method of discriminating between aggressive and non-aggressive people. As a result, this study had a high WoE A and C score for outcome measures.

A Dutch translation of the Reactive Proactive Questionnaire (RPQ) (Cima et al., 2013) was used in Van Bockstaele et al.'s (2020) study and this was adapted to a four-point Likert scale, as opposed to the five-point Likert scale that is typical, due to experimenter error. They also included a hostile attribution bias assessment, which

is a variant of the interpretation recognition task (Houtkamp, et al., 2017). The weaknesses associated with the outcome measures used in this study were balanced by other strengths in relation to the intervention content and delivery (outlined above), which led to an overall high WoE A and C for outcomes.

Hudley et al. (1993) used a variety of measures to assess aggression (see Table 5), and, whilst none of the outcome measures in this study had been standardised, the use of multiple measures allowed for triangulation of data, which gave this study a high WoE A rating. As standardised measures were not used in this study, this lowered the WOE C rating for outcomes to 2.

In Abdulmalik et al.'s (2016) study, the author created the following outcome measures for the purpose of the study: a teacher assessment of aggression (qualitative observation); the teacher rating of students' aggressive behaviour (TRAB); a self-rated aggression scale (SRAS); and a social cognition and attribution scale (SCAS). Abdulmalik et al. (2016) also asked teachers to complete the Strengths and Difficulties questionnaire (SDQ) (Goodman, 2001). The use of multiple measures allowed for triangulation of data, however, insufficient information was provided on intervention content and implementation therefore this study achieved a low WoE A rating. Despite this, as a standardised measure was used in this study and triangulation of data was possible, the study was given a WOE C rating of 3 for outcomes.

Findings

Outcome and Effect Size

As the focus of this review was on the effects of AR on aggression, only effect sizes on general measures of aggression were reported. A summary of each study's effect sizes along with their main findings are summarised in Table 7 with the effect size descriptors outlined in Table 8. Cohen's *d* was extracted from one study (Van Bockstaele et al., 2020), whilst partial eta squared was extracted from a further study (Vassilopoulos et al., 2015) and converted into Cohen's *d* using the Psychometrica calculator (Lenhard & Lenhard, 2016) to enable comparison. The effect size was not quoted in the remaining studies so it was calculated using the F-statistic of the interaction between pre- and post- scores and the AR intervention and control groups using the Campbell Collaboration Online calculator (Wilson, n.d.).

All studies reported a large effect size, suggesting school-based AR interventions reduce aggression in school-age populations. This is in line with previous research that demonstrated the positive effect of AR on aggression (Graham, 1997). As all five studies reported no significant differences between groups, and all of the studies presented with methodologically sound effect sizes for the specified population, a meta-analysis was carried out to assess the overall effect of AR on aggression.

Table 7: Summary of the effect size for each study

Study	Main Findings	Outcome measures	Sample Size	Effect Size (ES)	WoE D
Vassilopoulos et al. (2015)	Participants receiving CBM-I were more likely to have benign attributions in response to ambiguous social scenarios. In addition, aggressive behaviour scores reduced more in the AR intervention group than in the control group. Children who received AR also reported less anger and reported showing more self-control than those in the control group. The authors did not carry out follow-up assessments so the long-term effect of AR is unclear.	The Aggression Scale (aggression) (Orpinas & Frankowski, 2001)	34	0.94 Large	3 High
Hudley et al. (1998)	Participants receiving the Brain Power Program showed improvements in behaviour which were related to changes in their attributions. Improvements in aggression were twice that of those in the control group from pre- to post-intervention. Further treatment effects diminished over time but this was less for the AR group than the control group.	Social Skills Rating System (SSRS-T) (aggression) (Gresham & Elliot, 1990)	89	1.19 Large	3 High
Van Bockstaele et al. (2020)	The AR intervention led to a reduction in hostile attribution bias, which decreased reactive aggression but not proactive aggression when compared with the control group. No follow-up	Adapted Reactive Proactive Questionnaire (aggression) (Dutch translation) (Cima et al., 2013)	39	0.85 Large	3 High

	assessments were carried out so the long-term effect of AR is unclear.				
Hudley & Graham (1993)	Aggressive participants in the AR intervention group were less likely to presume hostile intent from peers in hypothetical and lab-based simulations of ambiguous provocation post-intervention. Further to this, participants were rated as less aggressive in teacher feedback following the interventions completion. The authors did not carry out any follow-up assessments so the long-term effect of AR is unclear.	Laboratory Analogue task (aggression) - designed by authors for purpose of this study Teacher Checklist – aggression subscale (aggression) (Coie, 1990) Formal Disciplinary Referrals (aggression) – school-based logs	44	0.80 Large	3 High
Abdulmalik et al. (2016)	The AR intervention group had significantly lower teacher rating of aggressive behaviour and lower ratings on the self-rated aggression scale, 1-week post intervention. The authors reported that this demonstrated that the incidence of aggressive behaviours in the AR group had decreased relative to the control group. The authors did not carry out any follow-up assessments so the long-term effect of AR is unclear.	Teacher assessment (aggression) qualitative observation Teacher-rated SDQ (aggression) (Goodman, 2001) Teacher Rating of Student's Aggressive Behaviour (TRAB) (aggression) –designed by authors for this study Self-rates Aggression Scale (SRAS) (aggression) - designed by authors for this study Social Cognition and Attribution Scale (SCAS) (attribution) - designed by authors for this study	37	1.11 Large	2 Medium

Table 8: Descriptors for Cohen’s d and Hedge’s g (Cohen, 1992)

Effect Size	Descriptor
0.8	Large
0.5	Medium
0.2	Small

Meta-analysis

The meta-analysis was conducted using a random-effects model in Meta-Essentials software (Suurmond et al., 2017). A random-effects model was used due to the heterogeneity across studies in terms of the participants and intervention, meaning that there may be a range of ‘true’ effect-sizes (Borenstein et al., 2010). The comparison group in the meta-analysis was the control group in each study. The control groups in each study did not experience any intervention and so provide a useful contrast to the intervention group. A summary of effect sizes and meta-analysis parameters is provided in Table 9 with a Forest plot included in Figure 2.

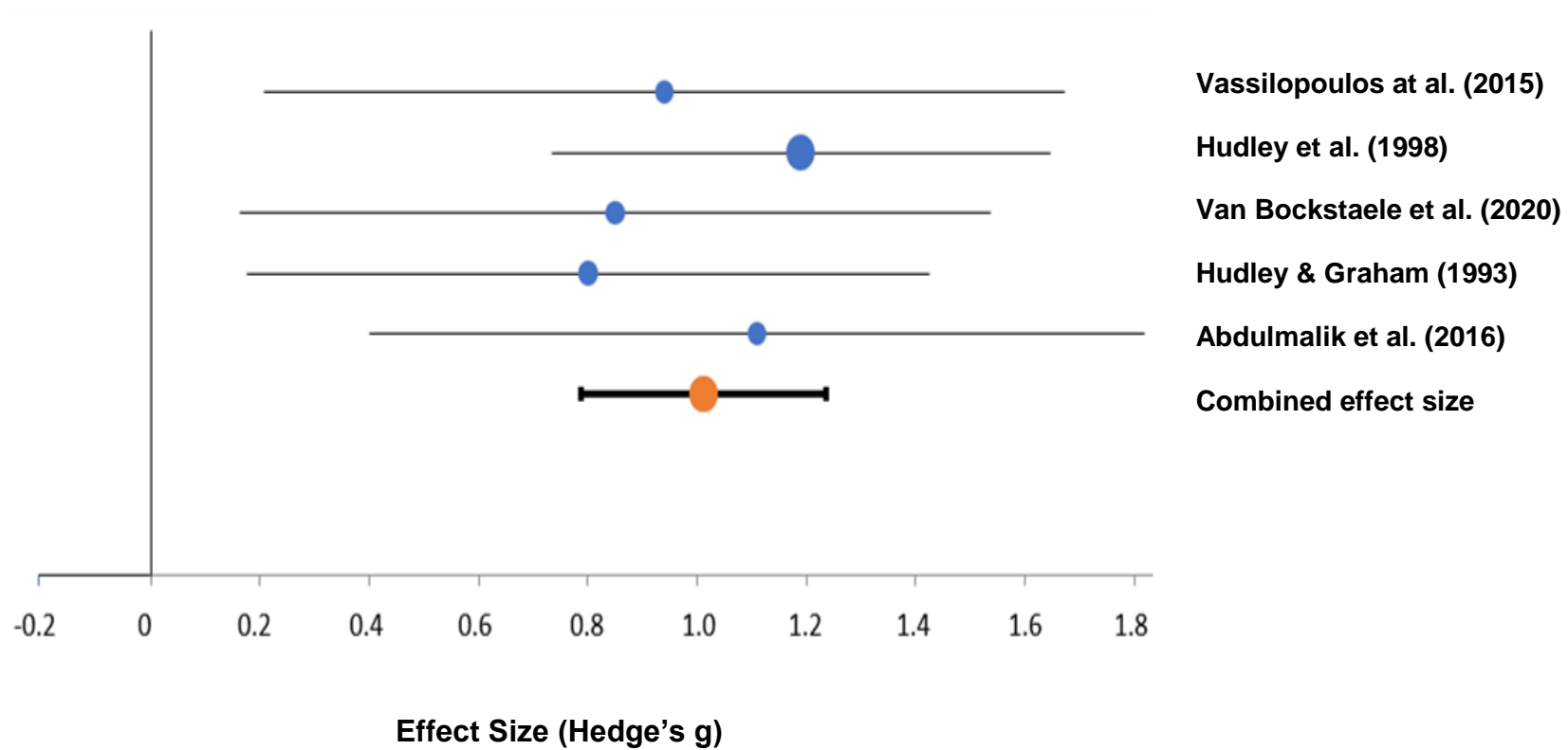
For the purposes of the meta-analysis, Cohen’s d was calculated using the pooled standard deviation from both groups and corrected for small sample upward-bias, yielding an overall effect across all studies of $g=1.01$ (95% $CI_{lower}=0.79$, $CI_{upper}=1.23$). This represents a large overall effect in favour of AR interventions, according to Cohen’s criteria. This is indicative of the large effect size that each study had. These correspond with high WoE D ratings in four of the studies with the exception being Abdulmalik et al. (2016), which received a medium WoE D. This means that more caution should be given when reporting the accuracy of this study’s results.

Table 9: Summary of effect sizes and meta-analysis parameters of the five studies included in this Meta-Analysis

Study	Treatment (N)	Control (N)	ES (SE)	95% CI	Weight
Vassilopoulos et al. (2015)	16	18	0.94 (0.36)	0.23; 1.65	14.23%
Hudley et al. (1998)	66	23	1.19 (0.23)	1.64; 0.74	35.26%
Van Bockstaele et al. (2020)	19	20	0.85 (0.34)	0.19; 1.50	16.62%
Hudley & Graham (1993)	20	24	0.80 (0.31)	0.18; 1.41	18.92%
Abdulmalik et al. (2016)	18	19	1.11 (0.35)	0.41; 1.80	14.97%
Combined	243		1.01 (0.32)	0.79; 1.23	

Note: ES – Effect Size; SE – Standard Error; CI – Confidence Intervals

Figure 2 Forest Plot of Effect Sizes (95% CI) of the Five Included Studies and the Combined Effect Size from the Meta-Analysis



As well as having a significantly large overall effect size, there was low heterogeneity in the findings (Cochran's $Q=1.41$, $p=0.842$, $I^2=0.00\%$, $t^2=0$). Accordingly, this indicates that the effect of the intervention is not highly variable between studies and so recommendations can be made based on the findings of this review.

Conclusions and Recommendations

Discussion of Findings

This review evaluated the effectiveness of school-based AR interventions in reducing peer-directed aggressive behaviour in children. Five studies met the inclusion criteria, with four receiving a high WoE D rating and one receiving a medium WoE D rating.

Combining the evidence of statistical effect, methodological quality, and methodological and topical relevance, it can be concluded that school-based AR interventions had a large effect on reducing peer-related aggression in aggressive children. Gersten et al. (2005) proposes that there needs to be at least two studies with high WoE A and a combined effect size significantly greater than zero for an intervention to be classified as 'evidence-based practice'. As four of the studies within this review had high WoE A ratings and the combined effect size was $g=1.01$, this review demonstrates an evidence base for the effectiveness of school-based AR interventions in reducing aggression in school aged populations.

It cannot, however, be concluded that the effects of AR are long-term as only one study assessed effectiveness 12 months post-intervention and the treatment effects had significantly reduced during this time, though to a lesser degree in the experimental group than the control group (Hudley et al. 1998). In addition, it cannot be concluded that this is the most effective intervention for reducing aggression. This

is because all of the studies had a no-intervention comparison group, meaning comparison between AR and other interventions was not possible.

Limitations of the Review

It could be argued that this review's inclusion criteria included studies which implemented AR interventions in a wide variety of formats, alongside utilising various outcome measures, which may have impacted the comparability of these studies. However, this was taken into consideration in WoE C and a random-effects meta-analysis was carried out in order to try to mitigate the impact of this.

The author of this review also developed their own protocol for WoE C, which was necessary due to the specificity of the review question and lack of existing protocols in this field. However, this could be questioned as it was not possible to test it for construct or inter-rater reliability prior to review submission.

Recommendations for Future Research

Given, the lack of UK-based research on the effectiveness of AR interventions on aggression in school-aged children, the extent to which this approach could be modified for the UK education system could be an area for future research.

In addition, the long-term benefits of AR interventions are unclear and were only investigated by one study. Therefore, further studies could include follow-up measures to evaluate maintenance and generalisability of effect.

Recommendations for Practice

EPs have an ethical responsibility to ensure that any interventions they recommend have a current and sound evidence base (BPS, 2018). As this review demonstrates an evidence base for school-based AR interventions, EPs may consider

recommending it to schools where there is a concern about aggressive behaviour. This is important given the damaging effect aggressive behaviour can have on the long- and short-term outcomes for children.

Despite this, none of the studies were carried out in the UK and therefore, it is unclear whether the results are generalisable to the UK school population, given its cultural disparity and the different schooling systems. This may impact upon EPs' confidence in recommending AR, without further UK-based evidence. In addition, the findings of this review suggest that the effectiveness of AR interventions decreased over time, which may give EPs further reservations.

Despite this, within the current review, results did replicate between different countries, languages and school systems. Therefore, it may be inferred that similar results would be found within a UK sample. In addition, only one study in this review conducted follow-up assessments and therefore it is not possible to make robust conclusions without further research on long-term impact.

Overall, given the benefits of school-based AR interventions for aggressive school children, it should be seriously considered for use by EPs, with a view that further research could be carried out to explore the UK context and longer-term effects.

References

- Abdulmalik, J., Ani, C., Ajuwon, A. J. & Omigbodun, O. (2016). Effects of problem-solving interventions on aggressive behaviours among primary school pupils in Ibadan, Nigeria. *Child and Adolescent Psychiatry and Mental Health*, 10 (31), 1-10. <https://doi.org/10.1186/s13034-016-0116-5>
- Barker, C., Pistrang, N. & Elliott, R. (2015). *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners*. Wiley-Blackwell.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111. <https://doi.org/10.1002/jrsm.12>
- BPS. (2018). *Code of Ethics and Conduct (2018) | BPS*. The British Psychological Society. Retrieved on 6th January 2021 from: <https://www.bps.org.uk/news-and-policy/bps-code-ethics-and-conduct>
- Chodkiewicz, A. R. & Boyle, C. A. I. (2016). Promoting positive learning in Australian students aged 10 to 12 years old using attribution retraining and cognitive behavioral therapy: A pilot study. *School Psychology International*, 37(5), 519–535. <https://doi.org/10.1177/0143034316667114>
- Cima, M., Raine, A., C. Meesters, C. & Popma, A. (2013). Validation of the Dutch Reactive Proactive Questionnaire (RPQ): Differential correlates of reactive and proactive aggression from childhood to adulthood. *Aggressive Behavior*, 39, 99-113.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Coie, J. (1990). *Teacher Checklist*. Unpublished manuscript.

- Crick, N. R. & Dodge, K. (1994). A review and reformulation of social information processing mechanism in children's social adjustment. *Psychological Bulletin*, 115 (1), 74-101.
- Dodge, K. & Coie, J. (1987). Social information processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology*, 53, 1146-1158.
- Dodge, K. (1980). Social Cognition and children's aggressive behavior. *Child Development*, 51, 167-170.
<https://doi.org/10.1111/j.1467-8624.1980.tb02522.x>
- Dodge, K. A., Pettit, G. S., McClaskey, C. L., Brown, M. M. & Gottman, J. M. (1986). Social Competence in Children. *Monographs of the Society for Research in Child Development*, 51, 1-85. <http://dx.doi.org/10.2307/1165906>
- Dodge, K. (2006). Translational science in action: Hostile attribution style and the development of aggressive behaviour problems. *Development and Psychopharmacology*, 18, 791-814.
- Frederickson, N. & Cline, T. (2007). *Special Educational Needs, Inclusion and Diversity: A Textbook*. Open University press: London
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C. & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71, 149–164.
<https://doi.org/10.1177/001440290507100202>
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337–1345.

- Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, 22 (2), 213–228.
<https://doi.org/10.1080/02671520701296189>
- Graham, S., Hudley, C. & Williams, E. (1992). Attributional and emotional determinants of aggression among African-American and Latino young adolescents. *Developmental Psychology*, 28, 731-74.
- Graham, S. (1997). Using attribution theory to understand social and academic motivation in African American Youth. *Educational Psychologist*, 32(1), 21-34.
- Gresham, F. M. & Elliott, S. N. (1990). *Social Skills Rating Scale*. Circle Pines: MN: American Guidance Service
- Harris M. B. & Huang, L. C. (1974). Aggression and the Attribution Process. *The Journal of Social Psychology*, 92 (2), 209-216.
<https://doi.org/10.1080/00224545.1974.9923100>
- Hawkins, K. A. and Cogle, J.R. (2013). Effects of interpretation training on hostile attribution bias and reactivity to interpersonal insult. *Behavior Therapy*, 44, 479-488.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley
- Hilt, L. M. (2004) Attribution Retraining for Therapeutic Change: Theory, Practice, and Future Directions. *Imagination, Cognition and Personality*, 23 (4), 289-307. <https://doi.org/10.2190/WC91-VXQK-N665-2J2Y>
- Houtkamp, E. S., van der Molen, M. J., Salemink, E., de Voogd E. L. & Klein A. M. (2017). Interpretation biases in socially anxious adolescents with a mild intellectual disability. *Research in Developmental Disabilities*, 67, 94-98.

- Hudley, C. & Graham, S. (1993). An Attributional Intervention to Reduce Peer-Directed Aggression among African-American Boys. *Child Development, 64* (1), 124-38.
- Hudley, C. (1994). The reduction of childhood aggression using The Brain Power Program. In M. Furlong & D. Smith (Eds.) *Anger, hostility and aggression: assessment, prevention and intervention strategies for youth* (pp. 313-344). New York: Wiley.
- Hudley, C., Britsch, B., Wakefield, W. D., Smith, T., Demorat, M. & Cho, S. J. (1998). An attribution retraining program to reduce aggression in elementary school students. *Psychology in the schools, 35* (3), 271-282.
- Lapointe, J. M. & Legault, F. (2004). Solving Group Discipline Problems without Coercion: An Approach Based on Attribution Retraining. *Journal of Classroom Interaction, 39* (1), 1-10.
- Lansford, J. E., Malone, P. S., Dodge, K. A., Pettit, G. S. & Bates, J. E. (2010). Developmental cascades of peer rejection, social information processing biases, and aggression during middle childhood. *Development and Psychopathology, 22*, 593–602. <https://doi.org/10.1017/s0954579410000301>
- Lenhard, W. & Lenhard, A. (2016). Calculation of Effect Sizes. Retrieved on 11th February from: https://www.psychometrica.de/effect_size.html. Dettelbach (Germany): Psychometrica. <https://doi.org/10.13140/RG.2.2.17823.92329>
- Lochman, J. E. & Dodge, K. A. (1998). Distorted perceptions in dyadic interactions of aggressive and nonaggressive boys: Effects of prior expectations, context, and boys' age. *Development and Psychopathology, 10*, 495–512. <https://doi.org/10.1017/S0954579498001710>.

- Lochman, J. E. & Wells, K. C. (2002). Contextual social cognitive mediators and child outcome: A test of the theoretical model in the Coping Power Program. *Development and Psychopathology*, 14, 945–967.
<https://doi.org/10.1017/s0954579402004157>
- Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal*, 339. <https://doi.org/10.1136/bmj.b2535>
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11 (2), 364–386.
<https://doi.org/10.1177/1094428106291059>
- Nasby, W., Hayden B. & Depaulo B. M. (1980). Attributional bias among aggressive boys to interpret unambiguous social-stimuli as displays of hostility. *Journal of Abnormal Psychology*, 89, 459-468.
- Novaco, R.W. (2003). *The Novaco Anger Scale and Provocation Inventory: Manual*. Western Psychological Services, Los Angeles.
- Orpinas, P. & Frankowski, R. (2001). The aggression scale: a self-report measure of aggressive behavior for young adolescents. *The Journal of Early Adolescence*, 21, 50–67.
- Petticrew, M. & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, 57, 527–529.
- Reed, R. (1988). Education and achievement of young, black males. In J. Gibbs (Ed.), *Young, black, and male in America: An endangered species* (pp. 37-96). Dover, MA: Auburn House

- Reef, J., Diamantopoulou, S., Vanmeurs, I., Verhulst, F. & vander Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behaviour and adult DSM-IV disorder: results of a 24-year longitudinal study. *Social Psychiatry Psychiatric Epidemiology*, 46 (12), 1233–41.
<https://doi.org/10.1007/s00127-010-0297-9>
- Robertson, J. S. (2000). Is attribution training a worthwhile classroom intervention for K–12 students with learning difficulties? *Educational Psychology Review*, 12 (1), 111-134.
- Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, 8 (4), 537–553.
<https://doi.org/10.1002/jrsm.1260>.
- Van Bockstaele, B., van der Molen, M. J., van Nieuwenhuijzen, M. & Salemink, E. (2020). Modification of hostile attribution bias reduces self-reported reactive aggressive behavior in adolescents. *Journal of Experimental Child Psychology*, 194.
- Vassilopoulos, S. P., Banerjee, R. & Prantzalou, C. (2009). Experimental modification of interpretation bias in socially anxious children: changes in interpretation, anticipated interpersonal anxiety, and social anxiety symptoms. *Behaviour Research and Therapy*, 47, 1085–1089.

- Vassilopoulos, S. P., Andreas Brouzos, A. & Andreou, E. (2015). A Multi-Session Attribution Modification Program for Children with Aggressive Behaviour: Changes in Attributions, Emotional Reaction Estimates, and Self-Reported Aggression. *Behavioural and Cognitive Psychotherapy*, 43 (5), 538 – 548.
<https://doi.org/10.1017/S1352465814000149>
- Verhoef, R. E. J., Alsem, S. C., Verhulp, E. E. & De Castro, B. O. (2019). Hostile Intent Attribution and Aggressive Behavior in Children revisited: A Meta-Analysis. *Child Development*, 90 (5), 525-547.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92 (4), 548-573.
- Wilmshurst, L. (2009). *Abnormal child psychology: A developmental perspective*. Routledge/Taylor & Francis Group.
- Wilson, D. B. (n.d.). *Campbell Collaboration Effect Size Calculator*. Retrieved January 5, 2021, from
<http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculatorSMD3.php>
- Ziegler, A. & Heller, K. A. (2000). Effects of an attribution retraining with female students gifted in physics. *Journal for the Education of the Gifted*, 23 (2), 217–243.

Appendix A

A list of the references for the studies, which were excluded at full text screening is provided in Table A. References are not provided for studies excluded at title and abstract screening but the total number of studies excluded is shown in Figure 1.

Table A: Articles excluded after full-text screening

Excluded Study	Rationale for exclusion	Exclusion criteria (Table 2)
Sukhodolsky, D. G., Golub, A., Stone, E. C. & Orban, L. (2005). Dismantling Anger Control Training for Children: A Randomized Pilot Study of Social Problem-Solving Versus Social Skills Training Components. <i>Behavior Therapy</i> , 36 (1), 15-23. https://doi.org/10.1016/S0005-7894(05)80050-4	Attribution Retraining (AR) is not a separate invention therefore it cannot be assessed whether AR leads to a reduction in aggression.	4
Hudley, C. & Friday, J. (1996). Attributional bias and reactive aggression. <i>American Journal of Preventive Medicine</i> , 12 (5), 75-81.	There are no results recorded in this study	9
Lapointe, J. M. & Legault, F. (2004). Solving Group Discipline Problems without Coercion: An Approach Based on Attribution Retraining. <i>Journal of Classroom Interaction</i> , 39 (1), 1-10.	This study aimed to reduce group discipline issues not aggression and involved teacher-student relationships not peer relationships	9
Hudley, C., Graham, S. & Taylor, A. (2007). Reducing Aggressive Behavior and Increasing Motivation in School: The Evolution of an Intervention to Strengthen School Adjustment. <i>Educational Psychologist</i> , 42, 251-260. https://doi.org/10.1080/00461520701621095	This is a report outlining how the Brainpower AR intervention came in to being	5

Hudley, C. A. (1992, April 20-24). <i>The Reduction of Peer Directed Aggression among Highly Aggressive African-American Boys</i> [Paper presentation]. American Educational Research Association, San Francisco, California, USA. https://eric.ed.gov/?id=ED346204	Not a peer-reviewed journal article	8
Hudley, C. A. (1994, April 4-6). <i>Attribution Retraining and Behavior Change among Highly Aggressive and Nonaggressive African-American Boys</i> [Paper presentation]. American Educational Research Association, New Orleans, LA, USA.	Not a peer-reviewed journal article	8
Hudley, C. A. (1995, March 30-April 2). <i>Reducing Peer Directed Aggression in the Elementary Grades: The Effects of an Attribution Retraining Program</i> [Paper presentation]. Meeting of the Society for Research in Child Development. Indianapolis, IN, USA.	Not a peer-reviewed journal article	8
Jon Hobbs, L., & Yan, Z. (2008). Cracking the walnut: Using a computer game to impact cognition, emotion, and behavior of highly aggressive fifth grade students. <i>Computers in Human Behavior</i> , 24 (2), 421-438. https://doi.org/10.1016/j.chb.2007.01.031	Attribution Retraining (AR) is not a separate invention	4
van Dijk, A., Thomaes, S., Poorthuis, A. M. G. & de Castro, B. (2019). Can self-persuasion reduce hostile attribution bias in young children? <i>Journal of Abnormal Child Psychology</i> , 47 (6), 989–1000. https://doi.org/10.1007/s10802-018-0499-2	This study did not research if AR leads to a reduction in aggression directly. It looked at it indirectly through self-persuasion.	9
Vassilopoulos S.P., Brouzos A. & Rentzios C. (2014). Evaluation of a universal social information-processing group program aimed at preventing anger and aggressive behaviour in primary school children. <i>Hellenic Journal of Psychology</i> , 11 (4), 208-222.	A quasi-experimental design.	5

Appendix B

Criteria and rationale for all Weight of Evidence (WoE) Ratings

WoE A: Methodological Quality

WoE A enables the user to judge whether a study has been well executed (Gough, 2007). The Gersten et al. (2005) coding protocol was used to analyse all five studies as it is designed to evaluate group experimental designs, such as the RCTs in this review. This coding protocol includes both essential and desirable criteria.

Adjustments were made to some of the questions within this protocol and are outlined in Appendix C with rationale. The criteria weighting for each WoE A rating is provided in Table B1 below. A summary of scores for each study is provided in Table B2. A completed example of a WoE A protocol is provided in Appendix D.

Table B1: Criteria for WoE Ratings

WoE A Rating	Criteria
3 High	Study meets at least 9 essential criteria and at least 5 desirable criteria
2 Medium	Study meets at least 9 essential criteria and at least 1 desirable criteria
1 Low	Study meets at least 6 but fewer than 9 essential criteria or does not meet any desirable criteria

Note: 'no evidence' would have been quantified by a zero rating but this did not apply to any studies in this review

Table B2: Each studies' summary of scores for WoE A

	Essential Criteria (/10)				Desirable Criteria (/10)	Overall WoE A Rating
	Participants (/3)	Intervention implementation and comparison conditions (/3)	Outcome measures (/2)	Data analysis (/2)		
Vassilopoulos at al. (2015)	2	3	2	2	6	3 High
Hudley et al. (1998)	3	3	2	2	4	3 High
Van Bockstaele et al. (2020)	2	3	2	2	6	3 High
Hudley & Graham (1993)	3	3	2	1	8	3 High
Abdulmalik et al. (2016)	2	2	2	2	5	1 Low

Note: WoE ratings are described as: 'High' = 3; 'Medium' = 2; and 'Low' = 1.

WoE B: Methodological Relevance

WoE B enables the quality and relevance of the research design of the studies in this review to be judged relative to the review question (Gough, 2007). For this review, WoE B considered the relevance of the methodology for evaluating the effectiveness of AR interventions in reducing peer-related aggression in school-based settings.

The coding protocol for WoE B was developed by the author and based on Petticrew and Roberts (2003) 'typology of evidence'. This enabled the author to analyse the strengths and weaknesses of each study's methodological approach relative to the question of 'effectiveness' in this review. Petticrew and Robert's (2003) findings on the type of studies most suited for 'effectiveness' questions and the criteria allocated for 'low (1), 'medium' (2) and 'high' (3) ratings are outlined in table B3 below along with the rating for each study. A completed example of a WoE B protocol is provided in Appendix D.

Table B3: Each studies WoE B Rating

	Vassilopoulos et al. (2015)	Hudley et al. (1998)	Van Bockstaele et al. (2020)	Hudley & Graham (1993)	Abdulmalik et al. (2016)
3 (High)					
Randomised Controlled trials (RCTs)	3	3	3	3	3
At least one control/comparison group					
2 (Medium)					
Cohort studies or Quasi-experimental studies (non-random assignment)					
At least one control/comparison group					
1 (Low)					
Research that collects qualitative data, surveys, non-experimental studies					
No control/comparison group					

Note: WoE ratings are described as: 'High' = 3; 'Medium' = 2; and 'Low' = 1.

WoE C: Topic Relevance

Weight of Evidence C (WoE C) evaluates the relevance of the focus of the study relative to the review question. Studies were rated on five areas: intervention, outcomes, setting, implementation and instructors, and the criteria are outlined in more detail below along with the rationale (see Table B4). The WoE C ratings for each study are outlined in Table B5 and a completed example of a WoE C protocol is provided in Appendix D.

Table B4 Criteria and Rationale for WoE C Ratings

Criteria	Rating	Rationale
A. Intervention	3- AR is the sole intervention or the primary feature of the main intervention. 2- AR is combined with another intervention or adapted in a minor way, which does not impact the fidelity of the AR aspect of the intervention being carried out. 1- Attribution retraining is not a core feature of the intervention.	It is important to understand how effective AR was in the intervention condition. As a result, any studies which do not have AR as the main component will be excluded.
B. Outcomes	3- Outcomes measured using a standardised assessment of aggression 2- Outcomes are not measured using a standardised assessment of aggression. 2- Multiple non-standardised measures such as an instructor constructed assessment of aggression and/or peer rating scale(s) have been used	Outcome measures must accurately assess the effect AR has on aggression. Standardised measures can provide a more accurate portrayal but use of multiple measures allows for triangulation which can improve the accuracy of results.

	1- Outcomes measures indirectly assess aggression e.g. hostile attribution bias, anger.	
C. Setting	<p>3 The intervention was conducted in a classroom setting within the school the participant is attending</p> <p>2- The intervention was carried out in a separate location within the school the participant is attending</p> <p>1- The intervention was carried out in a non-school setting e.g. clinic</p>	<p>In order for results to be generalisable, the ideal location for the intervention would be in the classroom. However, within the same school could still be beneficial in this respect. Studies which conducted AR interventions outside of school have been excluded as this review is looking at school-based interventions.</p>
D. Implementation	<p>3- The intervention is multi-faceted and involves a variation of instruction, practice, direct attributional feedback and consolidation work.</p> <p>2- The intervention provides direct attributional feedback only.</p> <p>1- The intervention provides indirect attributional feedback.</p>	<p>AR interventions can vary in terms of their implementation, however, the more rigorous and multi-faceted the approach the more likely it is to be effective.</p>
E. Instructor	<p>3- The intervention is delivered by an AR-trained member of staff at the school the participant is attending</p> <p>2- The intervention is delivered by an AR-trained researcher</p> <p>1- The intervention is delivered by an individual with minimal AR-training</p>	<p>It is thought the more familiar the instructor the more successful and generalisable the intervention is likely to be. In addition, it is thought that the more rigorous the training of the instructor, the greater the fidelity of the intervention implementation. If the instructor is untrained, the intervention is less likely to be effective in AR.</p>

Table B5: Each studies' summary of scores for WoE C

Study	Criteria Rating					Average WoE C
	A	B	C	D	E	
Vassilopoulos at al. (2015)	3 High	3 High	2 Medium	2 Medium	2 Medium	2 Medium
Hudley et al. (1998)	3 High	3 High	2 Medium	3 High	2 Medium	3 High
Van Bockstaele et al. (2020)	3 High	3 High	2 Medium	2 Medium	2 Medium	2 Medium
Hudley & Graham (1993)	3 High	2 Medium	2 Medium	3 High	3 High	3 High
Abdulmalik et al. (2016)	2 Medium	3 High	3 High	2 Medium	2 Medium	2 Medium

Note: WoE ratings are described as: 'High' = 3; 'Medium' = 2; and 'Low' = 1. Each Criteria Rating were given equal weighting and the sum of these were averaged and rounded to the nearest whole number to give an overall Weight of Evidence C (WoE C)

Weight of Evidence D (WoE D): Overall Weight of Evidence

WoE D provides a judgement of the overall weight of evidence for each study and is calculated by averaging the ratings from WoE A, B and C for each study. This overall rating is important as it specifies the extent to which each study has contributed evidence towards the review question. The WoE D for each study is outlined in Table B6.

Table B6: Weight of Evidence D for included studies

Study	WoE A	WoE B	WoE C	WoE D
Vassilopoulos et al. (2015)	3 High	3 High	2 Medium	3 High
Hudley et al. (1998)	3 High	3 High	3 High	3 High
Van Bockstaele et al. (2020)	3 High	3 High	2 Medium	3 High
Hudley & Graham (1993)	3 High	3 High	3 High	3 High
Abdulmalik et al. (2016)	1 Low	3 High	2 Medium	2 Medium

Note: WoE ratings are described as: 'High' = 3; 'Medium' = 2; and 'Low' = 1. WoE A, B and C were given equal weighting and the sum of these were averaged and rounded to the nearest whole number to give an overall Weight of Evidence D (WoE D).

Appendix C

Adaptations made to the WoE A coding protocol (Gersten et al., 2005)

Changes were made to Gersten et al.'s (2005) coding protocol to improve the clarity of the text and ensure it was appropriate for the review question posed. Deletions are shown by ~~strikes~~ through the text, additional text is in {curly brackets}, and the rationale for the changes are written in [square brackets]. All other text is the same as the original. As a result of these changes, there were ten desirable criteria rather than the original 8. The criteria of Gersten et al. (2005) for a 'high' quality study were altered to require 5, rather than 4, desirable criteria and this is shown in Table B1 above.

Essential Quality Indicators

Quality indicators for describing participants

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

[Rationale: 'difficulties' is more appropriate terminology than 'disabilities' in relation to the concept of aggression]

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

Was sufficient information given characterizing the ~~interventionists~~ {instructors} or teachers provided? Did it indicate whether they were comparable across conditions?

[Rationale: change of 'interventionists' to 'instructors' to ensure consistency in the language used throughout this review]

Quality indicators for implementation of the intervention and description of comparison conditions

Was the intervention clearly described and specified?

Was the fidelity of implementation described and assessed?

Was the nature of services provided in comparison conditions described?

Quality indicators for outcome measures

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

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

Quality indicators for data analysis

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?

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

Desirable Quality Indicators

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%?~~

{Was severe overall attrition (30% or more) avoided? Is attrition comparable across samples?}

[Rationale: It was possible that some studies may meet the first part of the criteria but would still be coded negatively if they did not meet the second part, therefore the questions were split into two parts to allow for each to be graded separately (see {...} above for additional question)]

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?~~

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

[Rationale: It was possible that some studies may meet the first part of the criteria but would still be coded negatively if they did not meet the second part, therefore the questions were split into two parts to allow for each to be graded separately (see {...} above for additional question)]

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

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

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?

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

Did the research report include actual {written,} audio or videotape excerpts that capture the nature of the intervention?

[Rationale: It would be possible for a study to capture the nature of the intervention in written form with similar levels of ecological validity to audio or videotape excerpts therefore it was deemed that this should be included as a further option]

Were results presented in a clear, coherent fashion?

Overall Rating of Evidence: 3 2 1 0

Appendix D

Example of a completed WoE A coding protocol from one study

Coding protocol:

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71, 149-164.

Study 1: Vassilopoulos et al. (2015)

Essential Quality Indicators

Quality indicators for describing participants

Was sufficient information provided to determine whether the participants demonstrated the difficulties presented?

Yes

No

Unknown/Unable to Code

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

Yes

No

Unknown/Unable to Code

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

Yes

No

Unknown/Unable to Code

Quality indicators for implementation of the intervention and description of comparison conditions

Was the intervention clearly described and specified?

Yes

No

Unknown/Unable to Code

Was the fidelity of implementation described and assessed?

- Yes
- No
- Unknown/Unable to Code

Was the nature of services provided in comparison conditions described?

- Yes – test-retest – no intervention condition
- No
- Unknown/Unable to Code

Quality indicators for outcome measures

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

- Yes
- No
- Unknown/Unable to Code

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

- Yes – pre and post assessment
- No
- Unknown/Unable to Code

Quality indicators for data analysis

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

- Yes
- No
- Unknown/Unable to Code

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

- Yes – partial eta squared = .18 equates to Cohen's d = 0.94
- No
- Unknown/Unable to Code

Desirable Quality Indicators

Was data available on attrition rates among intervention samples?

Yes – assume no one left the study based on N reported in results being same as in methods section

No

Unknown/Unable to Code

Was severe overall attrition (30% or more) avoided? Is attrition comparable across samples?

Yes

No

Unknown/Unable to Code

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

Yes

No

Unknown/Unable to Code

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

Yes

No

Unknown/Unable to Code

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

Yes

No

Unknown/Unable to Code

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

Yes

No

Unknown/Unable to Code

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

- Yes
- No
- Unknown/Unable to Code

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

- Yes – no intervention control group – test-retest
- No
- Unknown/Unable to Code

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

- Yes – written examples of materials used
- No
- Unknown/Unable to Code

Were results presented in a clear, coherent fashion?

- Yes
- No
- Unknown/Unable to Code

Overall Rating of Evidence: 3 2 1 0

Example of a completed WoE B coding protocol from one study

Coding protocol:

Created for the purposes of this review. Criteria rationale based on “Typology of evidence” recommendations for research best suited to studying the effectiveness of interventions (Petticrew and Roberts, 2003).

Study 1: Vassilopoulos et al. (2015)

Criteria WoE B Rating Criteria 3

Randomised Controlled trials (RCTs)

At least one control/comparison group

Criteria WoE B Rating Criteria 2

Cohort studies or Quasi-experimental studies (non-random assignment):

At least one control/comparison group

Criteria WoE B Rating Criteria 1

Research that collects qualitative data, surveys, non-experimental studies:

No control/comparison group

Overall Rating of Evidence: 3 2 1 0

Example of a completed WoE C coding protocol from one study

Coding protocol:

Created for the purposes of this review.

Study 1: Vassilopoulos et al. (2015)

A. Intervention

3- Attribution retraining is the primary intervention or the core feature of the main intervention condition.

2- Attribution retraining is combined with another intervention.

1- Attribution retraining is the secondary or tertiary intervention.

As many studies combine attribution retraining with other interventions, it is important to understand how significant of a feature attribution retraining was in the intervention condition.

B. Outcomes

3- Outcomes have been measured using a standardised assessment of aggression.

2- Outcomes have been measured using a teacher/instructor constructed assessment of aggression and peer rating scale.

1- Outcomes have been measured using an assessment that indirectly measures aggression.

This question focuses on the effect of attribution retraining on aggression, so outcome measures must accurately measure this. Standardised and/or validated measures will produce a more accurate portrayal.

C. Setting

3 The intervention was conducted in a classroom setting.

2- The intervention was carried out in a separate location within a school.

1- The intervention was carried out in a non-school setting such as a psychology lab or office.

As the intervention is intended to be used in school, the study should also take place in a realistic school setting in order for results to be generalisable.

D. Implementation

3- The intervention involves direct attributional feedback, instruction, practice and consolidation.

2- The intervention involves direct attributional feedback.

1- The intervention involves indirect attributional feedback through modelling, face-to-face or via video.

Attribution retraining does not have clear implementation guidelines but it does contain some key elements that should be included. The level of which these features are included will impact how closely the study reflects how attribution retraining will be implemented in practice.

E. Instructor

3- The intervention is delivered by a member of the teaching staff who has been trained in Attribution Retraining.

2- The intervention is delivered by a researcher who has been trained in Attribution Retraining.

1- The intervention is delivered by an individual who has received minimal training in Attribution Retraining.

The intervention is intended to be delivered in schools and by trained school staff, so if the study does so it will more closely reflect how Attribution Retraining will be implemented in practice. If the instructor is untrained in Attribution Retraining then the intervention is unlikely to be effective in retraining attributions.

Overall Rating of Evidence: 3 2 1 0